An Ethics of Artificial Intelligence Curriculum for Middle School Students

Blakeley H. Payne, blakeley@mit.edu

MIT Media Lab

Personal Robots Group directed by Cynthia Breazeal

August 2019

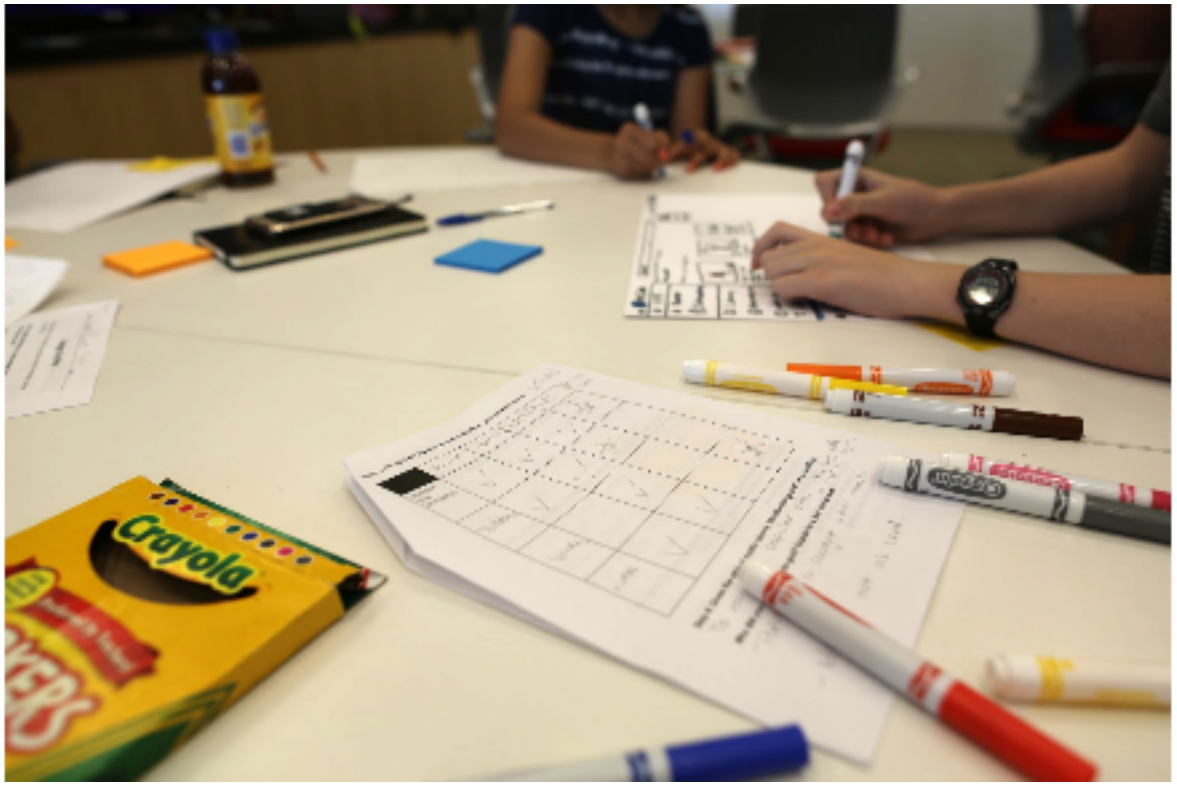
A student shows off her paper prototype for her redesign of YouTube.

Please consider giving us feedback at:

https://mit.co1.qualtrics.com/jfe/form/SV_6X5UWiD7p58BnNz

Table of Contents

Introduction



A pair of students work on their paper prototype of YouTube after completing an ethical matrix.

About

This document includes a set of activities, teacher guides, assessments, materials, and more to assist educators in teaching about the ethics of artificial intelligence. These activities were developed at the MIT Media Lab to meet a growing need for children to understand artificial intelligence, its impact on society, and how they might shape the future of AI.

This curriculum was designed and tested for middle school students (approximately grades 5th-8th). Most activities are unplugged and only require the materials included in this document, although unplugged

modifications are suggested for the activities which require computer access.

Usage

License: CC-BY-NC under Creative Commons

These materials are licensed as CC-BY-NC under creative commons. This license allows you to remix, tweak, and build upon these materials non-commercially as long as you include acknowledgement to the creators. Derivative works should include acknowledgement but do not have to be licensed as CC-BY-NC.

To acknowledge the creators, please include the text, "An Ethics of Artificial Intelligence Curriculum for Middle School Students was created by Blakeley H. Payne with support from the MIT Media Lab Personal Robots Group, directed by Cynthia Breazeal."

More information about the license can be found at: https://creativecommons.org/licenses/by-nc/4.0/

People interested in using this work for for-profit commercial purposes should reach out to Cynthia Breazeal at cynthiab@media.mit.edu for information as to how to proceed.

How to Access Materials

In order to use and edit the materials below, please make a copy of this document by:

1. Making sure you are logged into your Google account.
2. Go to File > Make a copy

3. You will then be prompted to name and save the materials to your drive.

All slides can be found at:
https://drive.google.com/open?id=1gp2Hywu8sOoweEc2NQb6V-yiXAuSolnC
Each slide deck is also linked below with its corresponding activity. In order to access slides, make sure to follow the steps above to add them to your Google Drive.

Translated Materials

We are very fortunate that these materials have been translated by many individuals all across the globe. You can access these translated materials here:

German: https://thingminds.ch/de/kikids/

Translated by Eugen Rodel of thingminds.

Portuguese (partial translation as well as remixed materials):
https://drive.google.com/drive/folders/1NhCTVNm-qg5BaMAdiTjvG1Ulebh6OJqE

Translated by Miguel Angelo Abreu Sousa of the Federal Institute of Sao Paulo, Brazil.

Korean:
https://keris.or.kr/main/ad/pblcte/selectPblcteETCInfo.do?mi=1142&pblcteSeq=13255

Translated by Dr. Han-Sung Kim of the Korea Education and Research Information Service.

Feedback

Thank you for checking out our AI + Ethics Curriculum! We plan to continuously evaluate and iterate on this work, so please consider filling out the following survey to give us your feedback (~5 min in length):

https://mit.co1.qualtrics.com/jfe/form/SV_6X5UWiD7p58BnNz

Additional feedback you may have to prg.ai.ethics@gmail.com. We'd love to hear which resources you use, with what age groups, and any feedback you'd like to give for future iterations of this curriculum!

Learning Objectives

1. **Understand the basic mechanics of artificial intelligence systems.**
    a. Recognize algorithms in the world and be able to give examples of computer algorithms and algorithms in everyday contexts (for example, baking a cake).
    b. Know three parts of an algorithm: input, steps to change input, output.
    c. Know that artificial intelligence is a specific type of algorithm and has three specific parts: dataset, learning algorithm, and prediction.
        i. Understand the problem of classification in the supervised machine learning context.
        ii. Understand how the quantity of training data affects the accuracy and robustness of a supervised machine learning model.
    d. Recognize AI systems in everyday life and be able to reason about the prediction an AI system makes and the potential datasets the AI system uses.
2. **Understand that all technical systems are socio-technical systems. Understand that socio-technical systems are not neutral sources of information and serve political agendas.**
    a. Understand the term "optimization" and recognize that humans decide the goals of the socio-technical systems they create.

b. Reason about the goals of socio-technical systems in everyday life and distinguish advertised goals from true goals (for example, the YouTube recommendation algorithm aims to make profit for the company, while it is advertised as a way to entertain users).
    i. Map features in existing socio-technical systems to identified goals.
c. Know the term "algorithmic bias" in the classification context.
    i. Understand the effect training data has on the accuracy of a machine learning system.
    ii. Recognize that humans have agency in curating training datasets.
    iii. Understand how the composition of training data affects `the outcome of a supervised machine learning system.

3. **Recognize there are many stakeholders in a given socio-technical system and that the system can affect these stakeholders differentially.**
    a. Identify relevant stakeholders in an socio-technical system.
    b. Justify why an individual stakeholder is concerned about the outcome of a socio-technical system.
    c. Identify values an individual stakeholder has in an socio-technical system, e.g. explain what goals the system should hold in order to meet the needs of a user.
    d. Construct an ethical matrix around a socio-technical system.

4. **Apply both technical understanding of AI and knowledge of stakeholders in order to determine a just goal for a socio-technical system.**
    a. Analyze an ethical matrix and leverage analysis to consider new goals for a socio-technical system.
    b. Identify dataset(s) needed to train an AI system to achieve said goal.
    c. Design features that reflect the identified goal of the socio-technical system or reflect the stakeholder's values.
5. **Consider the impact of technology on the world.**
    a. Reason about secondary and tertiary effects of a technology's existence and the circumstances the technology creates for various stakeholders.

Activities

Students work together to build their paper prototypes of YouTube.

Overview

The following table provides an overview of the activities included in the curriculum:

Name	Description	Standards	Time
AI Bingo	Students are given bingo cards with various AI systems. Students find a partner who has also used that AI system and together work to identify what prediction the system is making and the dataset it uses.	1.c, 1.d	30 min
Algorithms as	Students learn that algorithms, like recipes,	1.b	45

Opinions	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.	2.a, 2.c	min
Ethical Matrix	Building on the algorithms as opinions lesson, students identify the stakeholders who care about their peanut butter and jelly sandwich algorithm and the values those stakeholders have in the algorithm. They then fill out an ethical matrix to see where those values overlap or conflict.	2.b 3.a, 3.b, 3.c, 3.d 4.a	45 min
Intro To Supervised Machine Learning & Algorithmic Bias*	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. When the classifier works better on cats than dogs, students have the opportunity to retrain their classifiers with their own new datasets.	1.c 2.c	~ 3 hrs
Speculative Fiction**	Students have the opportunity to interact with various technologies, such as emotion detection software or GANs. Students then respond to creative writing prompts about who might be affected by the technology and how the technology might produce harm or benefit in the future. This activity is not unplugged, but instructions for an unplugged version are included in the open source materials.	3.a, 3.b, 3.c 5.a	~3 hrs
YouTube Scavenger Hunt	Similar to AI Bingo, students in partners are tasked to recognize the various AI systems on the YouTube platform (e.g., advertisement matching algorithm, the recommender algorithm, comment classifier, etc) For each system, students identify what the algorithm is trying to predict and the dataset the algorithm uses.	1.d 2.b	30 min

YouTube Redesign	Students apply what they have learned so far by constructing an ethical matrix around the YouTube Recommender Algorithm. Based on this ethical matrix, students determine a goal (or "opinion") for their algorithm. Students then paper prototype what this new version of YouTube would look like and imagine features that meet the values their identified stakeholders have.	2.b 3.a, 3.b, 3.c, 3.d 4.a, 4.b, 4.c	~4 hrs
YouTube Socratic Seminar	Students read an abridged version of a Wall Street Journal article titled *YouTube Weighs Major Changes to Kids Content Amid FTC Probe* and then participated in a socratic seminar discussing which stakeholders were most important or influential to the proposed changes to the YouTube Kids app and whether or not technologies like autoplay should exist.	2.b 3.a, 3.b, 3.c 4.c 5.a	30 min

*Denotes the activity is not unplugged.

**Denotes the activity is not unplugged but includes instructions about how to modify it to be unplugged.

AI Bingo

Description

Description	Learning Objectives	Time
Students are given bingo cards with various AI systems. Students find a partner who has also used that AI system and together work to identify what prediction the system is making and the dataset it uses.	1.c, 1.d	30 min

Slides

Slides can be found at:
https://docs.google.com/presentation/d/1oWLJmM9piLf4dGbmwQj5heRAnLulR98OSUZNXQyHQRI/edit?usp=sharing

Instructions

After reviewing the introductory slides, pass out bingo cards. Students must find a partner who has used an AI system listed on the card and together students must identify the prediction the system is trying to make and the dataset it might use to make that prediction. The first student to get five squares filled out in a row, diagonal, or column wins (or, for longer play, the first student to get two rows/diagonals/columns).

After playing, have students discuss which the squares they filled out.

Worksheet

On next page...

A.I. BINGO

Gotten a weather forecast from a website or used a weather app Dataset: Prediction:	Sent a voice-to-text message Dataset: Prediction:	Used an online search engine like Google or Bing Dataset: Prediction:	Seen a Google autofill search result Dataset: Prediction:	Had a writing assignment graded by a computer Dataset: Prediction:
Used “safe search” on Google Dataset: Prediction:	Seen a suggested response on Gmail to an email Dataset: Prediction:	Used a Snapchat filter (what’s your favorite?) Dataset: Prediction:	Played a motion-sensitive video game e.g. Mario Party, Nintendo, Wii U, etc. Dataset: Prediction:	Had an Emoji suggested instead of a word e.g. “lol” is replaced for an Emoji smiley face Dataset: Prediction:
Seen a sponsored product on Google or Amazon e.g. *”since you bought __, we thought you might like...”* Dataset: Prediction:	Had an email go to your spam folder (was it actually spam?) Dataset: Prediction:	**FREE**	Clicked on an Instagram ad (what kinds of ads do you normally see on the app compared to your partner?) Dataset: Prediction:	Seen news articles suggested in a news app (what kinds of articles do you normally see compared to your partner?) Dataset: Prediction:
Had an email labeled as “important” Dataset: Prediction:	Seen a suggested ad on Snapchat (if so, what for? How does this compare to what ads your partner sees?) Dataset: Prediction:	Had a text auto-completed or used autocorrect Dataset: Prediction:	Listened to a recommended song on Spotify (what kind of music do you usually get recommended compared to your partner?) Dataset: Prediction:	Seen a recommended product on Facebook (if so, what for?) Dataset: Prediction:
Seen a “nudge” reminder on Gmail to respond to an email Dataset: Prediction:	Used a fingerprint to unlock a device or opened a device with your face Dataset: Prediction:	Used a map app to find a path to a destination Dataset: Prediction:	Used an app to recognize a song playing Dataset: Prediction:	Communicated with a customer service bot Dataset: Prediction:

Introduction to Algorithms As Opinions

Description

Description	Learning Objectives	Time
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.	1.b 2.a, 2.c	45 min

Slides:

Slides can be found at:
https://docs.google.com/presentation/d/1wjwbwplu8uAWbHoQHs-B91vUVn5-LUoJ6lTjOtLq2RU/edit?usp=sharing

Teacher Guide

Teaching point: for students to understand that algorithms can have various goals and motives

Materials Required:

- Slides
- Worksheet

Connection: *Remember this morning when we learned what an algorithm is? Who can remind me what the three parts of an algorithm is?*

[wait for students to say input/data, specific steps to change that data, and an output]

That's correct. An algorithm needs some input data and follows specific steps or instructions to give us a desired output. Computers use algorithms, but so do humans. Algorithms are a lot like a recipe.

For example, if I were baking a cake, my algorithm would take in the following ingredients, like flour, sugar, salt, eggs, etc.

I would mix together my dry ingredients and then mix in the wet ingredients like eggs or milk.

I would pour into a cake pan, set the oven to 350, and put the cake pan in the oven.

My output would be a cake!

[show accompanying slides with images of cake]

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.

[let time elapse for students to work]

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.

[give students a few minutes to chat]

Who can share with the class what their algorithms had in common?

What was different?

If you have to give your algorithm a title "How to make the ____ PBJ," what adjective would you use? You can't use "best."

[most students will say yummiest/tastiest]

Possible questions:

1. *Did any of you include instructions to put away your ingredients after you used them?*
    a. *Then you were optimizing for tidiness in your algorithm!*
2. *Did any of you cut your sandwich into fun shapes? Cut off the crust?*
    a. *Then you were optimizing for playfulness or aesthetics!*

Computer algorithms also optimize for various goals, but sometimes this can be hard to spot. What do you think the goal of Google's search algorithm is?

[Students might say "best" results. If so, ask them what word they would replace with best like they did earlier. Students might also say "best results for me," so you can prompt students to ask what they mean by that, or how Google might confirm that they've shown "the best results for me." We're looking for answers like: to get us to click on links, to get us to click on advertisers links - things that show students understand the search results *benefit Google first*]

Optional: If possible, open up Google search under two different accounts (or one under an account that is logged in, and one in an incognito browser). Search for some of the following items: pizza place, best movie, news. Ask students why they think the results are different.

Name: _____ Date: _____

PB&J Sandwich Activity Sheet

Write an "algorithm" to make the <u>BEST</u> Peanut Butter and Jelly sandwich:

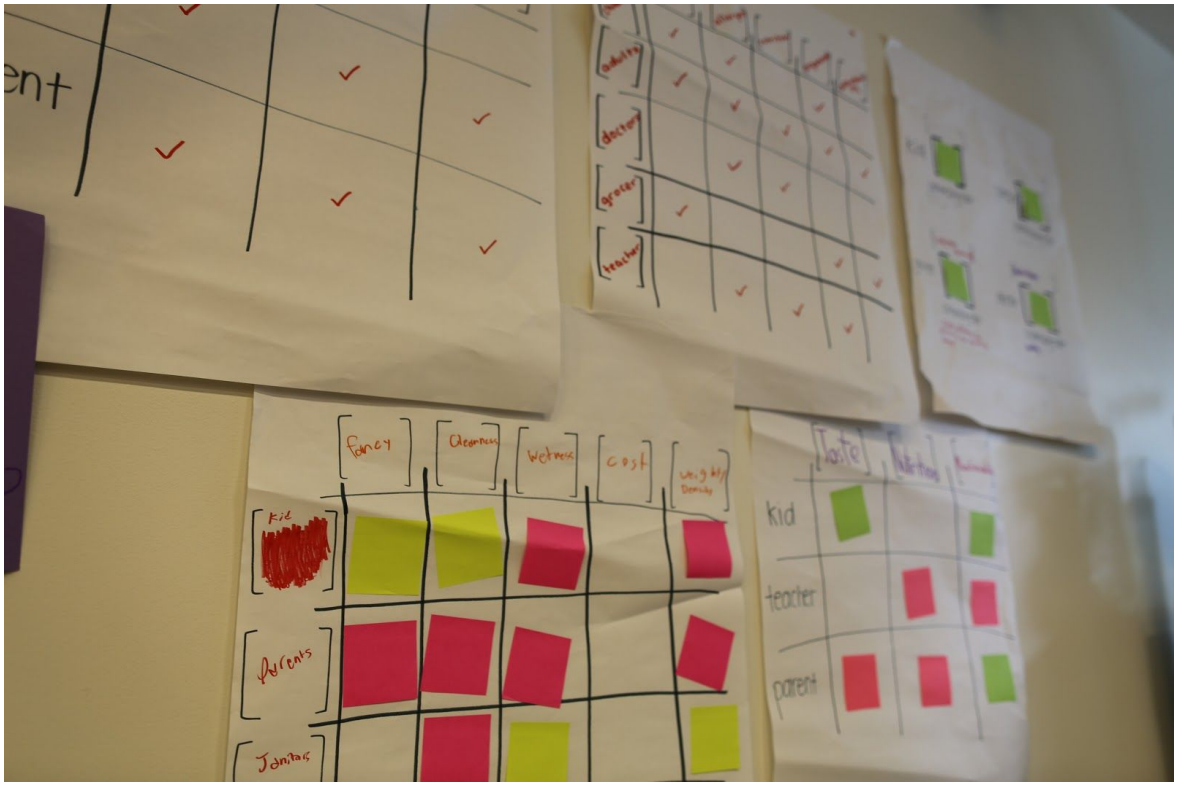
a) **What *input data* (or ingredients!) do you need?**

b) **Write out the steps in your algorithm:**

Ethical Matrix

Description

Description	Learning Objectives	Time
Building on the algorithms as opinions lesson, students identify the stakeholders who care about their peanut butter and jelly sandwich algorithm and the values those stakeholders have in the algorithm. They then fill out an ethical matrix to see where those values overlap or conflict.	2.b 3.a, 3.b, 3.c, 3.d 4.a	45 min



Example of students' ethical matrices for a peanut butter and jelly sandwich algorithm. Students used giant post-it notes instead of the worksheet below.

Slides

Slides can be found at:
https://docs.google.com/presentation/d/1AKGyEyNOxLI68y_sg3IIDxi_caWaDkTFpy-dtQLx28M/edit?usp=sharing

Teacher Guide

Teaching point: for students to be able to recognize stakeholders in a system and be able to build and utilize an ethical matrix

Materials Required:

- Butcher paper with ethical matrices drawn on and sticky notes *or* worksheet with relevant matrices printed out
- Ethical matrix worksheet
- Slide deck

Connection: *We've show that algorithms can have different goals. What were the goals of your PB&J algorithm?*

[pause for students to respond with answers like "tastiest," "healthiest," etc]

I want you to take five minutes with a partner to brainstorm some goals for your algorithm and write them down on sticky notes. Try and identify at least ten.

[Give students time to brainstorm. Answers could be quick, tasty, pretty/aesthetic, healthy/nutritious, cost-effective/cheap, easy to make]

Sharing out: *Which possible goals did you identify?*

Explicit Instruction: *Now, we are going to learn how we should decide what the goal of an algorithm should be.*

With your partner, I want you to brainstorm 5-10 different people who might care about your PB&J algorithm. Who cares about what you eat? Who cares about what you pack for lunch at school?

These people who care about your algorithms are called stakeholders. For each stakeholder, please list a reason or two why those stakeholders might care.

[Give students time to work in groups. Expect answers like: me, my parents, my siblings, my grandparents, my doctor/dentist, teacher, classmate, babysitter, sports coach.

To push students to think of more indirect stakeholders, ask them to think about their food supply. *Where do you get your food? Who makes money off of what food you buy? Do you think your grocer cares about your algorithm?* (Yes - the kind of peanut butter, bread, etc, are data inputs in the algorithm.) More indirect stakeholders may include: grocers, farmers, truckers, the companies who produce the peanut butter, jelly, etc]

Sharing out: *Which stakeholders did you come up with? Why do they care?*

Now I want to introduce you to the ethical matrix. On one side you will list your stakeholders and on another side you will list what they care about in your algorithm.

Explicit Instruction: Watch me as I fill out an ethical matrix for sandwich. **[demonstrate in slides]**

Now it's your turn. With your partner, fill out the ethical matrices in front of you. For each matrix, I want you to discuss what the goal of the algorithm should be for these stakeholders.

[give students time to fill out matrices]

Sharing out:

[for each matrix, ask students to justify why they crossed off each square and what the goal they decided for their algorithm]

Optional discussion questions:

How useful is the 1x1 ethical matrix?

[If students say yes, ask if it was useful for stakeholders other than themselves - e.g. if parents or doctors are the only stakeholders. Ask if they think this is fair.]

Did deciding a goal for your algorithm get easier or harder as the number of stakeholders went up? What about as the number of values changed?

What would help in deciding the goal of your algorithm?

[Ideas: Perhaps having stakeholders in the room to discuss, getting testimony from stakeholders]

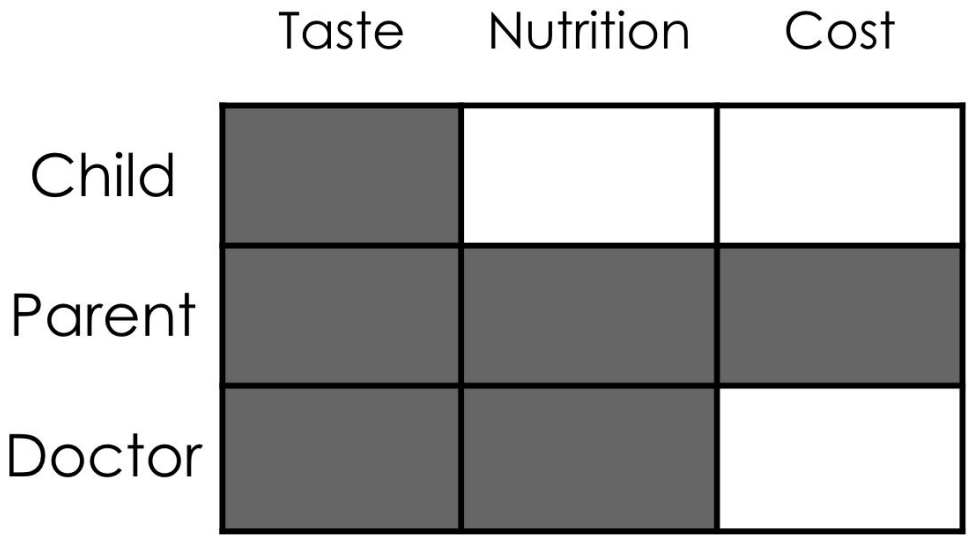
Name: _____ Date: _____

Ethical Matrix Activity Sheet

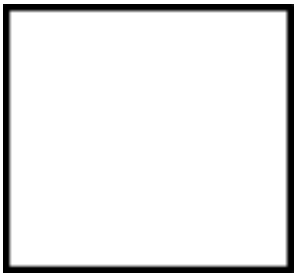
Instructions

Fill out the following ethical matrices. For each row, list a stakeholder who cares about the outcome of your algorithm. For each column, list a value a stakeholder may want to see expressed in your algorithm.

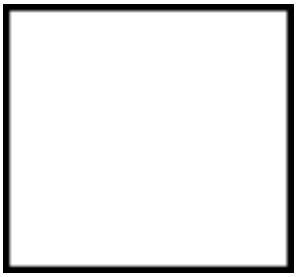
Example:



1x1 Ethical Matrix

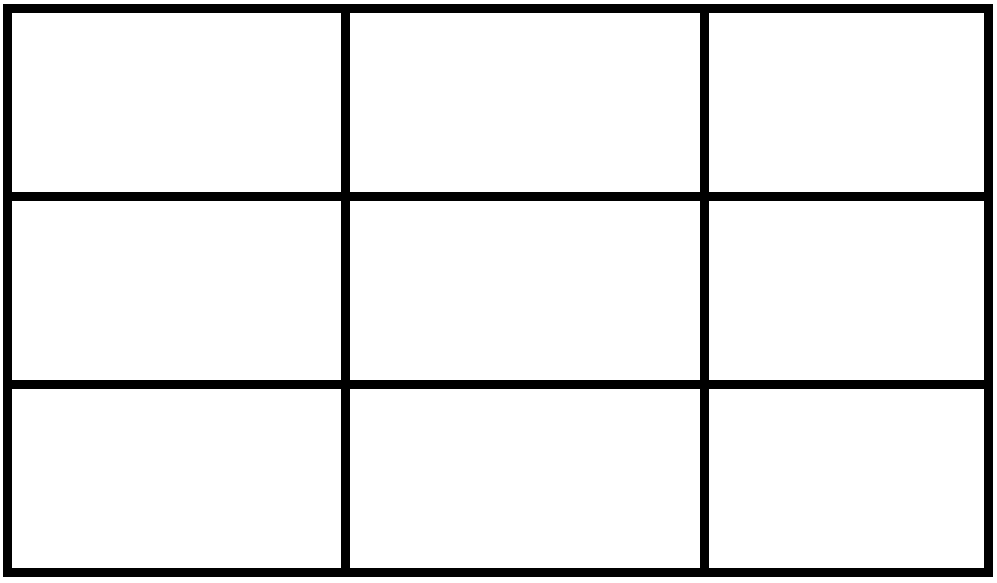


Goal:



Goal:

3x3 Ethical Matrix

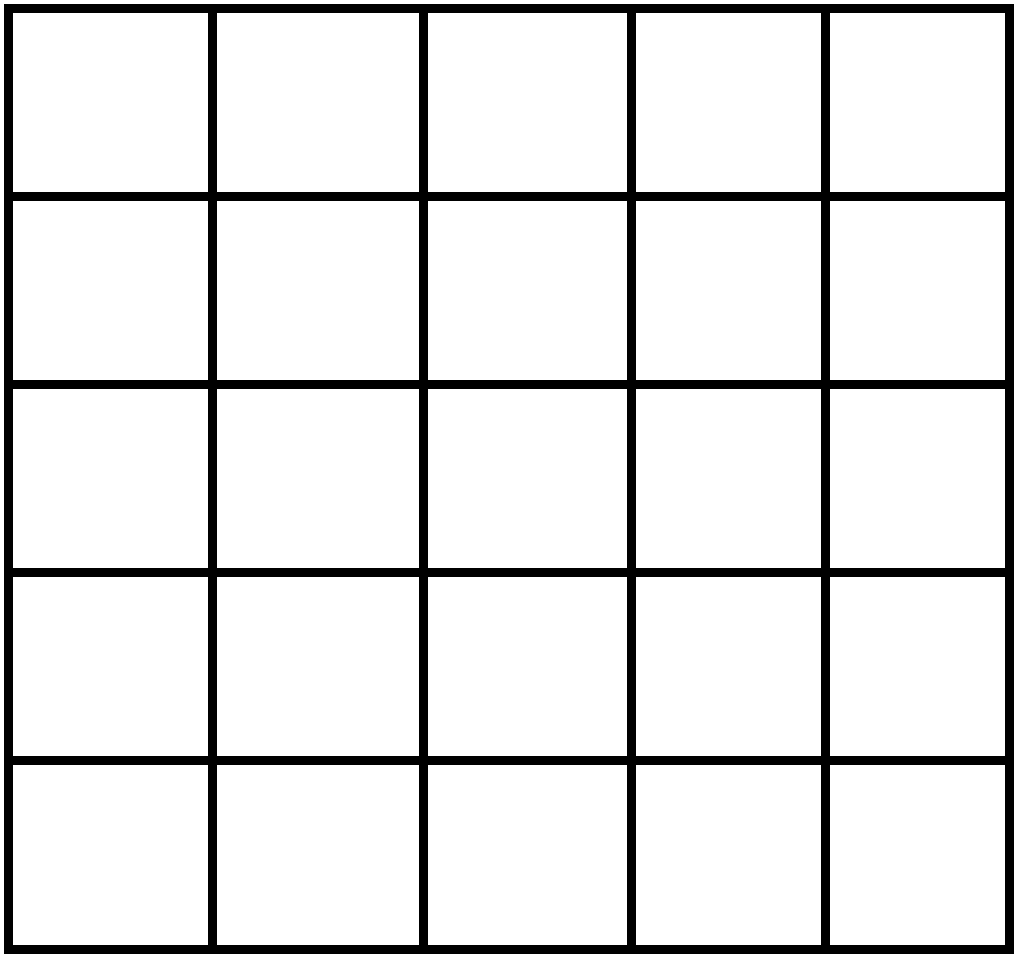


Goal:

2x3 Ethical Matrix

Goal:

5x5 Ethical Matrix



Goal:

Introduction to Supervised Machine Learning and Algorithmic Bias

Description

Description	Learning Objectives	Time
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. When the classifier works better on cats than dogs, students have the opportunity to retrain their classifiers with their own new datasets.	1.c 2.c	~ 3 hrs

Slides

Slides about supervised machine learning be found at:
https://docs.google.com/presentation/d/17mRdJxZFWYyzJiOmprcNfz3x2CWF1oPgn-8Rb4v3IVU/edit?usp=sharing

Slides about algorithmic bias can be found at:
https://docs.google.com/presentation/d/1DohsXiUt96pD42RSjJXHenu_AFInDNIzRMxWwXlKobs/edit?usp=sharing

Teacher Guide

Teaching point: For students to understand classification problems, and understand the role training data plays in classification accuracy

Materials Required:

- Slides
- Worksheet x 3
- Chromebooks + chargers
- Dog/cat cards
- Musical instrument packs

Connection: Remember yesterday when we discussed the definition of AI? What were the three components we talked about?

[wait for students to say dataset, learning algorithm, prediction]

Great! 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 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 by these labeled examples.

We're going to focus on classification problems today, but regression problems are very similar. With regression, instead of trying to predict a category a new piece of data belongs to, you are trying to predict some

numerical value for that data. For example, you might be trying to predict what the temperature will be tomorrow.

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.

[demo the Teachable Machines without the tutorial]

**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 with your partner. Write what you find on the activity sheets we gave you. 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?

Great. Now we are going to build a cat-dog classifier using the *new* version of teachable machines.

[demo how to train at least one class]

With your partner, I want you to build a machine that classifies cats and dogs. You already have training sets loaded on your laptops and can use the cards and webcam to test with.

[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., d*ogs 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?]*

Now we are going to build one more classifier. I'm going to give each of you a bag of musical instruments. I want you to build a classifier that classifies the instrument when it is being played.

Now, it's your turn. After you've gone through the training process, I want you to test your classifier playing multiple instruments and see what happens.

[give time to build classifier until students are clearly testing]

What happened when you played multiple instruments at once to test? Did it ever predict both instruments?

(No, because we've only taught it to classify one instrument at a time)

Sometimes classification can be really hard because it can be hard to list all the possible categories - think about how many classes we would need in order to classify every combination of musical instrument you have *(at least 24 classes)*.

This problem also showed up in Joy's work. Remember Joy's video? The biggest problem with the facial recognition systems she studied is that they don't work as well on darker, female faces as paler, male faces. But not everyone identifies as male or female, and those facial recognition systems can't capture that. This is something we always need to be careful of when we are classifying - to make sure that our classes don't exclude anyone.

Name: _____ Date: _____

Introduction to Supervised Machine Learning Activity

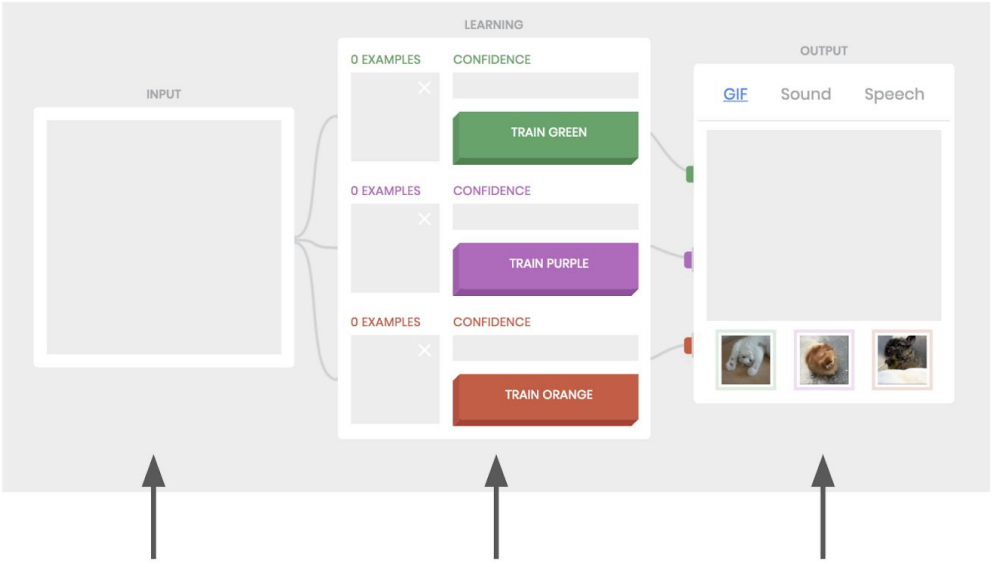
Description

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?

Image Datasets

Below are a set of images provided to students in order for them to build their cat-dog classifiers. Three different datasets are included:

Dataset	Description
Initial Training Dataset	These are the images students should use to "teach" their machine learning model which image is a cat and which image is a dog. Note that there are many more cats and that the cats are more diverse in appearance than the dogs. This means that the classifier will more accurately classify cats than dogs.
Test Dataset	These are the images that students should use to test their classifier after training. Students should show these images to their model and record if their classifier predicts if the image is of a dog or a cat. Note: Students should not use these images to teach their classifier. If an image is used to train a classifier, the machine will have already recorded the corresponding label for the particular image. Showing this image to the machine during the testing phase will not measure how well the model generalizes.
Recurating dataset	This is a large assortment of images students can use to make their training dataset of cats and dogs larger and more diverse.

The test dataset should be used twice, once for testing students' initial classifier and again for testing their recurated dataset.

The images are arranged below in a way that should make printing them off and cutting relatively easy (with two images per page, organized by dataset). Printing images allows students to hold them up in front of a webcam to train.

Initial Training Dataset

Dogs

Cats

Test Dataset

Recurating Dataset

Dogs

Cats

Name: _____ Date: _____

Introduction to Algorithmic Bias Activity Sheet

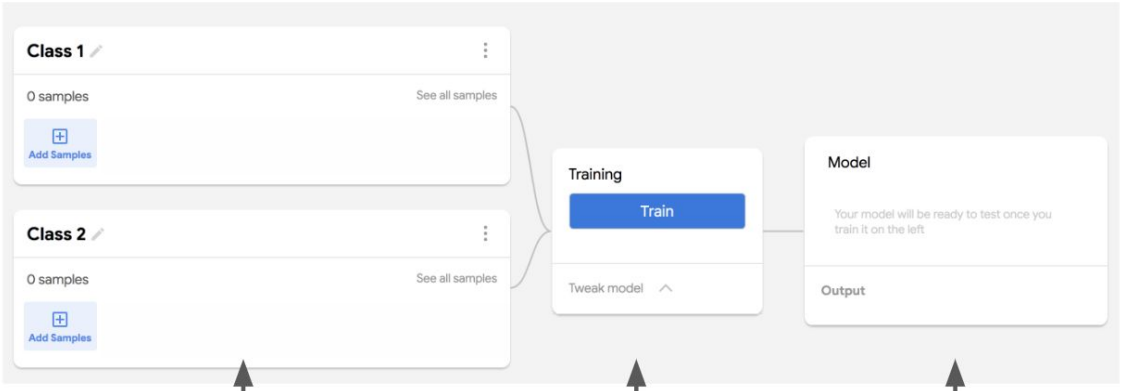
Description

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 "Train models on images."

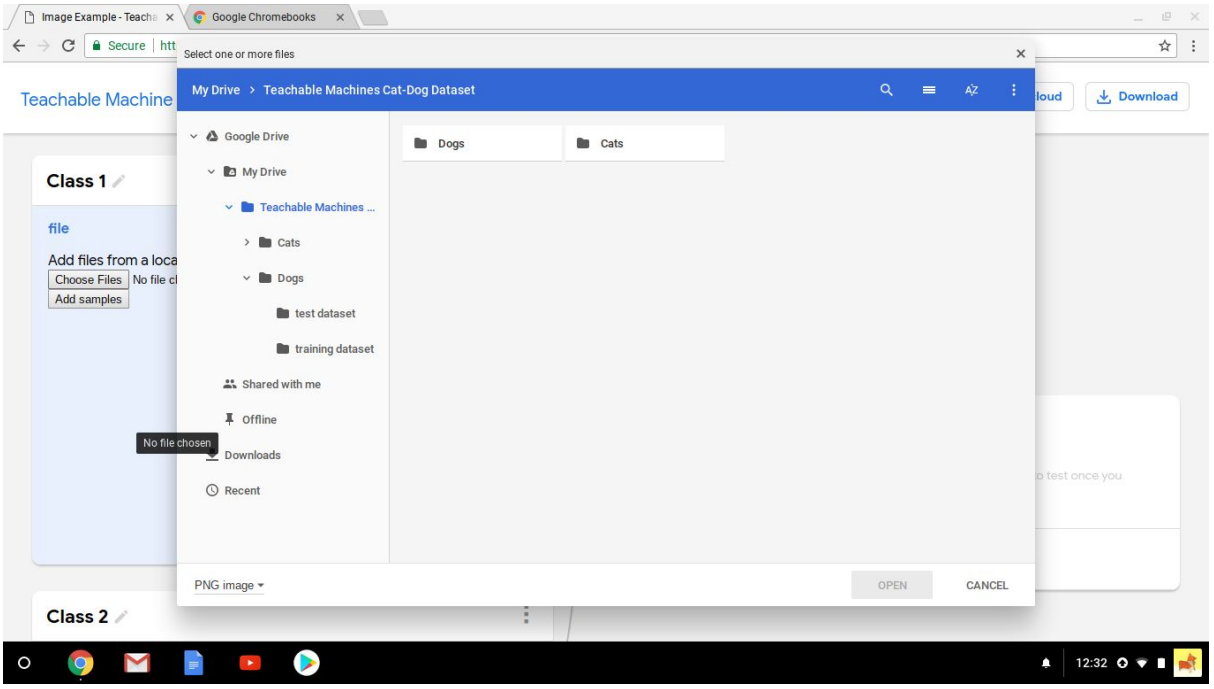
Question 1: Identify the three parts of an AI system in the teachable machine as discussed in class:



_____ _____ _____

Steps to add training datasets:

1. Click “Add samples”
2. Click “file”
3. Click “Choose files”
4. Locate the Teachable Machines Cat-Dog Dataset in the Drive:



5. Click on “Dogs” and then “training dataset”
6. Click “open” to add an image to the training set

Be sure to add all images <u>one time</u>.

7. After all of the images have been added, click “Add samples.”
8. Repeat with cat training images.

Question 2:

(a) For the dog training dataset, record the following:

How many images are included?

How are the images similar?

How are the images different?

(b) For the cat training dataset, record the following:

How many images are included?

How are the images similar?

How are the images different?

Question 3: **Train your classifier on the two training datasets.**

Once your classifier is finished, test your dataset with cards given to you containing the following image. Fill in the table on the next page about your testing dataset:

Image	Classification	Confidence Score	Correct?
			
			
			
			
			

Question 4: **Which class did your classifier work better on?** ***(Circle one)***

Cats **Dogs**

Why do you think that is?

Question 5: **With your group, use the photos on the tables to re-curate your training dataset. Record the following:**

A. **For the dog training dataset, record the following:**
   a. **How many images are included?**

   b. **How are the images similar?**

   c. **How are the images different?**

B. **For the cat training dataset, record the following:**

   a. **How many images are included?**

   b. **How are the images similar?**

   c. **How are the images different?**

Question 6: Train your new classifier on your two new training datasets.

Once your classifier is finished, test your dataset with cards given to you containing the following image. Fill in the table on the next page about your testing dataset:

Image	Classification	Confidence Score	Correct?
			
			
			

<u>Question 7:</u> **Did your new algorithm work… *(circle one)***

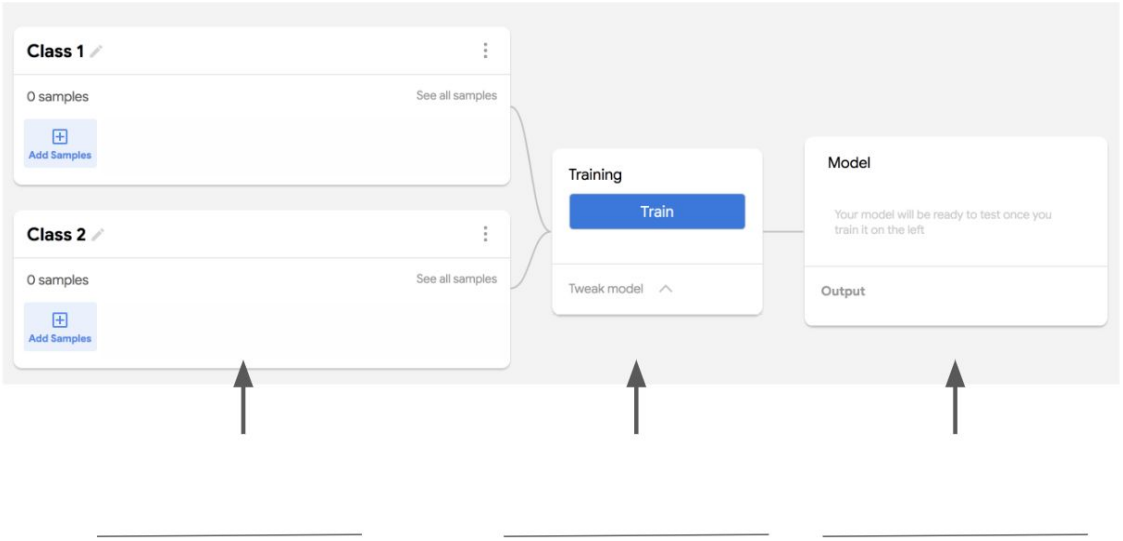
Better for dogs **The same for both cats and dogs** **Better for cats**

<u>Question 8:</u> **Explain your answer to Question 6:**

Name: _____ Date: _____

Supervised Machine Learning Quiz

1. What are the three components of an AI system?



_____ _____ _____

2. A supervised machine learning algorithm has been trained on the following images with the label "cat".

Image	 
Label	Cat

How will it classify the following image?

Circle one:

Cat **Dog**

3. A supervised machine learning algorithm has been trained on the following images:

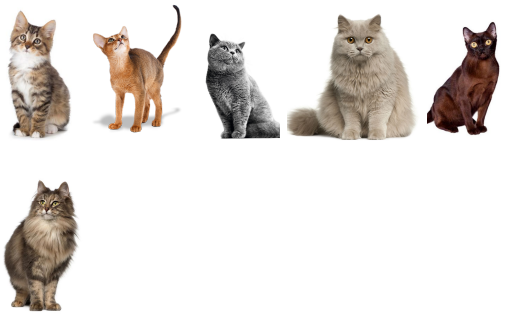
Image		
Label	Cat	Dog

3.1 Do you expect the algorithm's accuracy to be (circle one):

Better for cats **The same between cats and dogs** **Better for dogs**

3.2. Which training set, if any, would provide the best classification accuracy for both cats and dogs?

A.

	
Cat	Dog

B.

	
Cat	Dog

C.

	
Cat	Dog

D.


Dog

Speculative Fiction

Description

Description	Learning Objectives	Time
Students have the opportunity to interact with various technologies, such as emotion detection software or GANs. Students then respond to creative writing prompts about who might be affected by the technology and how the technology might produce harm or benefit in the future. This activity is not unplugged, but instructions for an unplugged version are included in the open source materials.	3.a, 3.b, 3.c 5.a	~3 hrs

Slides

Instructor slides can be found at:
https://docs.google.com/presentation/d/12p5keurZwQE137EuqkviKs3Iw506rmHWvNSpvRvscXk/edit?usp=sharing

Template slides for student presentations can be found at:
https://docs.google.com/presentation/d/1ATbbpyanbxT93rq9BiEQbKvEvohIhjcjNL2h4go8KG0/edit?usp=sharing

Instructions

Show students videos of Alter Ego (included in slides). After the video plays, ask students to summarize the technology (e.g. *"Can someone explain to me what Alter Ego does?"*).

Then lead students in a conversation about the ethical ramifications of the technology. Ask questions like:

- How could this technology be helpful? *(If students need prompting, ask them to identify examples from the videos)*
    - Who could this technology help?
    - What hopes do you have for this technology?
- How could this technology do harm?
    - How could it be bad?
    - What concerns do you have about this technology?
    - Who could this technology help?

After the discussion ends, tell students that they will have an opportunity to interact with a cutting-edge AI system and that they will give a presentation to the class about the technology.

Demo the Alter Ego presentation (in the slide deck) to the students and tell them their presentation will be similar. Show them where the can access their own template of the slides.

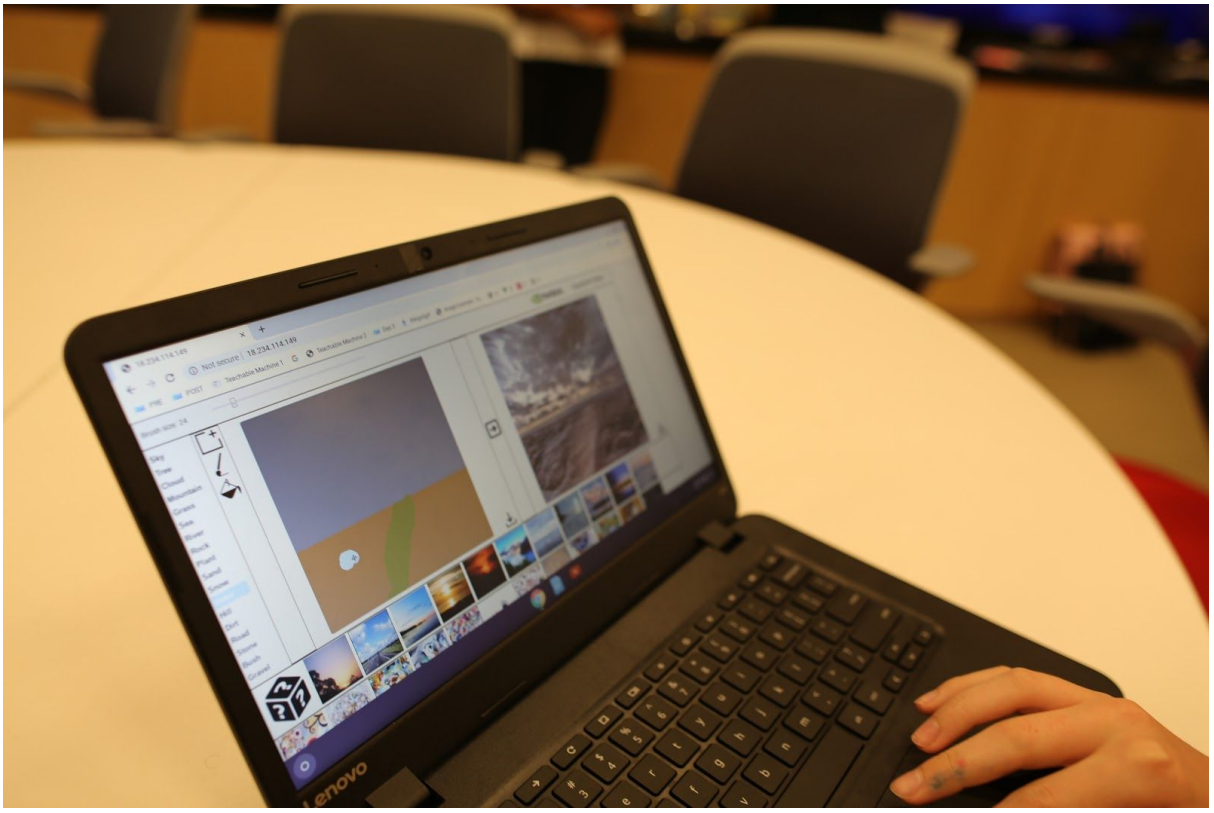
Divide students into partners or small groups and pass out the following worksheet. Have students play with the technology for ten minutes and then have them work together to fill out the prompts.

After completing the worksheet, students should make a copy of the slides and edit them to be about their technology. Then have all students present their slides to the class. Encourage students to ask speculative questions of the type *"And then what would happen?"* or *"What might be a ripple effect of that outcome?"* to their peers.

Unplugged Modification

In order to make this exercise unplugged, a teacher can do one of the following things:

- Show videos (like the Alter Ego video) and have students write in response to these videos
- Provide students with a description of the following technologies
- Have students, in groups, imagine a technology they think everyone should have in the future. Have groups explain their technology to another group and then each group answers the following questions about their partner group's technology.



A student plays with a GAN paint tool as part of the Speculative Fiction activity.

Name: _____ Date: _____

Speculative Futures Activity Version #1

Description

In this exercise, you will learn to think about the kind of world we make when we build new technology, and the unintended consequences that can occur when we build that technology.

Instructions

1. Go to: http://gandissect.res.ibm.com/ganpaint.html
2. Explore with the tool for al little bit!
3. Then, answer the following prompts:

Write a brief description of your technology:

Which stakeholders might be interested in this technology? Who might be affected by this technology most? Brainstorm at least 10 stakeholders.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

If this technology was used for evil, how might that be done?

If this technology was used to help other people, who might it help?

In 50 years this technology could do the most <u>good</u> by…

1.

2.

3.

In 50 years this technology could do the most <u>harm</u> by…

1.

2.

3.

After completing these prompts, raise your hand and ask for a Chromebook. WIth your partner, make a presentation about your technology.

Name: _____ Date: _____

Speculative Futures Activity Version #2

Description

In this exercise, you will learn to think about the kind of world we make when we build new technology, and the unintended consequences that can occur when we build that technology.

Instructions

1. Go to: https://talktotransformer.com/
2. Explore with the tool for al little bit!
3. Then, answer the following prompts:

Write a brief description of your technology:

Which stakeholders might be interested in this technology? Who might be affected by this technology most? Brainstorm at least 10 stakeholders.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

If this technology was used for evil, how might that be done?

If this technology was used to help other people, who might it help?

In 50 years this technology could do the most <u>good</u> by…

1.

2.

3.

In 50 years this technology could do the most <u>harm</u> by…

1.

2.

3.

Name: _____ Date: _____

Speculative Futures Activity Version #3

Description

In this exercise, you will learn to think about the kind of world we make when we build new technology, and the unintended consequences that can occur when we build that technology.

Instructions

1. Go to: https://demo.mr.affectiva.com/
2. Explore with the tool for al little bit!
3. Then, answer the following prompts:

Write a brief description of your technology:

Which stakeholders might be interested in this technology? Who might be affected by this technology most? Brainstorm at least 10 stakeholders.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

If this technology was used for evil, how might that be done?

If this technology was used to help other people, who might it help?

In 50 years this technology could do the most <u>good</u> by…

1.

2.

3.

In 50 years this technology could do the most <u>harm</u> by…

1.

2.

3.

Name: _____ Date: _____

Speculative Futures Activity Version #4

Description

In this exercise, you will learn to think about the kind of world we make when we build new technology, and the unintended consequences that can occur when we build that technology.

Instructions

1. Go to: http://deepangel.media.mit.edu/
2. Explore with the tool for al little bit!
3. Then, answer the following prompts:

Write a brief description of your technology:

Which stakeholders might be interested in this technology? Who might be affected by this technology most? Brainstorm at least 10 stakeholders.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

If this technology was used for evil, how might that be done?

If this technology was used to help other people, who might it help?

In 50 years this technology could do the most <u>good</u> by…

1.

2.

3.

In 50 years this technology could do the most <u>harm</u> by…

1.

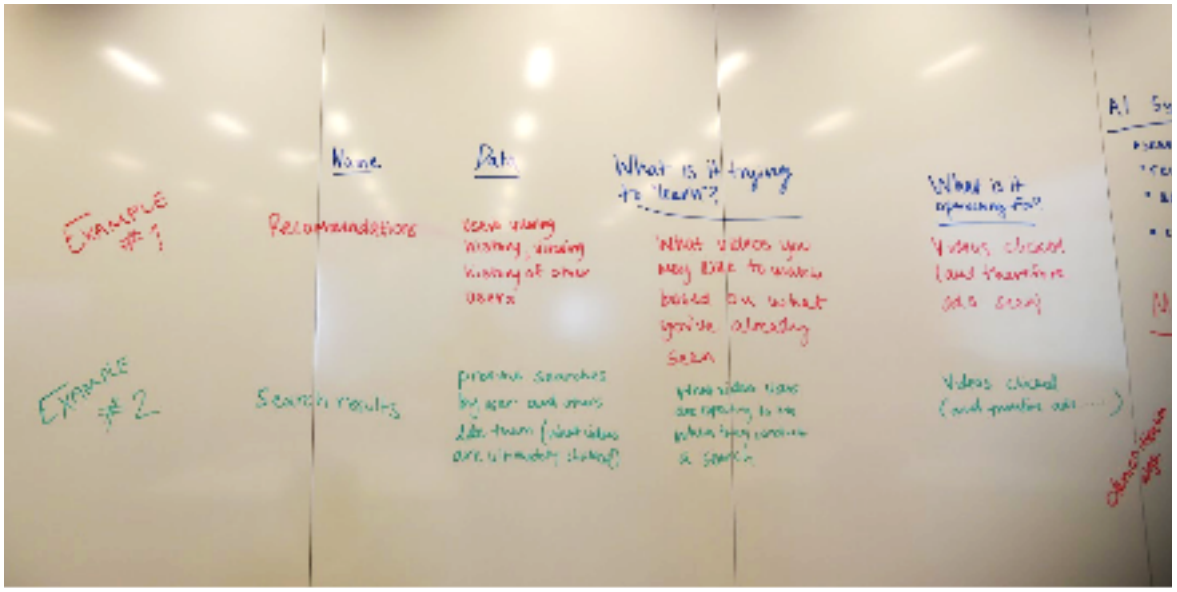
2.

3.

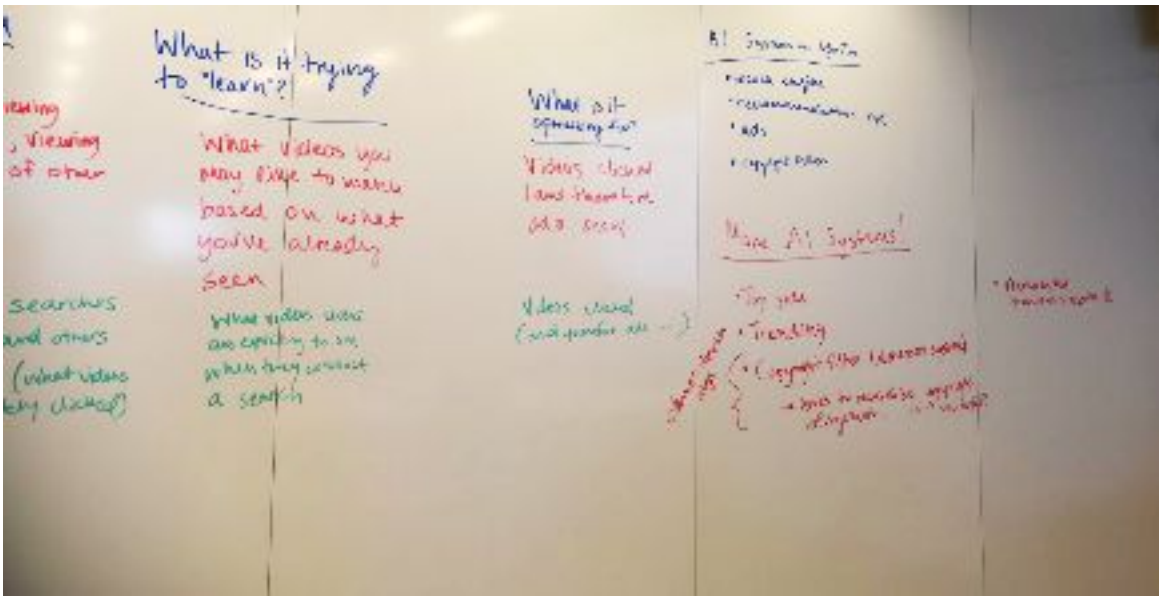
YouTube Scavenger Hunt

Description

Description	Learning Objectives	Time
Similar to AI Bingo, students in partners are tasked to recognize the various AI systems on the YouTube platform (e.g., advertisement matching algorithm, the recommender algorithm, comment classifier, etc) For each system, students identify what the algorithm is trying to predict and the dataset the algorithm uses.	1.d 2.b	30 min



Examples of answers and discussion for YouTube Scavenger Hunt



Examples of more answers and discussion for YouTube Scavenger Hunt

Teacher Guide

Teaching point: for students to understand that algorithms can have various goals and motives

Materials Required:

- Chromebooks
- Worksheet

Connection: We've just introduced you to your final projects where you will be redesigning YouTube. Before we begin brainstorming what our projects will look like, let's spend some time investigating the algorithms that YouTube is already using.

Break off into groups of 2-3 students and spend some time filling out the worksheet. You will uncover the ways in which YouTube uses AI. For each distinct feature, (1) describe the data that is being used to create the algorithm, (2) what the algorithm is trying to learn, (3) the overall "opinion" of the algorithm

[hand out worksheets and chromebooks]

Ok, so let's do this first one together. Open up your chromebooks to YouTube and start typing in the first 3 letters of your favorite type of dog. What happens?

[students say that YouTube fills in the rest of their search for them]

Yes, that's called a suggested search. What did you notice about the suggestions YouTube gave you?

[students share observations]

Try typing in a few more letters. What do you notice?

[students share observations]

What do you think the algorithm here is trying to learn?

[students say that it's trying to guess what you are searching]

Yes. It's trying to understand what you are looking for before you finish typing it in. What data do you think the algorithm is using to get there?

[students say previous searches by you and other people]

Yes, that's right. YouTube is using data from your previous searches as well as others like you. What do you think they are optimizing for?

[students say more videos watched so there is more ad revenue]

Great, so now I want you to work in partners to continue to search around google for other algorithms it may be using. Think about what the algorithm is trying to learn, what dataset it is using to learn that, and what it might be optimizing for. Do your best and raise your hands if you have any questions.

[students work for ~30 min]

Let's spend some time talking about what we found. Would anyone like to share an example they wrote down?

[begin writing examples down on the board. Looking specifically for the following]

Name	**Data**	**What is it trying to "Learn"?**	**What is it optimizing for?**
Suggested search (autofill)	Previous search history by you and others	What the most common searches are, so that it can predict what you want to see	More videos watched
Recommendations	User's viewing history, viewing history of other users	What videos you may like to watch based on what you have already seen	Videos clicked (and therefore ads seen)
Comments section	Amount of likes on the comment, popularity of those who commented	Which comments are the most relevant	Which comments will keep the user on the platform longer

Search results	Previous searches by user and others like them (what videos are ultimately clicked)	What videos users are expecting to see when they conduct a search	Videos clicked (and therefore ads…)
Ads	User's viewing history	What products and/or services might appeal to a user	Ad conversions
Autoplay	User's viewing history, viewing history of other users	What video a user will want to watch next	Videos watched; keep users on website

[once all algorithms are put on the board, ask the students questions about the process]

How did you go about figuring out what data was used to influence the algorithm? How did you decide on the opinion?

Was this task easy or difficult?
Is there anything that surprised you?

Name: _____ Date: _____

YouTube Scavenger Hunt Activity Sheet

Work with a partner to uncover the ways in which YouTube uses AI. For each distinct feature, (1) describe the data that is being used to create the algorithm, (2) what the algorithm is trying to learn, (3) the overall "opinion" of the algorithm.

Name	**Data**	**What is it trying to "Learn"?**	**What is it optimizing for?**

YouTube Redesign

Description

Description	Learning Objectives	Time
Students apply what they have learned so far by constructing an ethical matrix around the YouTube Recommender Algorithm. Based on this ethical matrix, students determine a goal (or "opinion") for their algorithm. Students then paper prototype what this new version of YouTube would look like and imagine features that meet the values their identified stakeholders have.	2.b 3.a, 3.b, 3.c, 3.d 4.a, 4.b, 4.c	~4 hrs

Slides

Slides can be found at:

https://docs.google.com/presentation/d/1zH8C1xQjLQrNaPdZJdAc7qW8y8UVsqeDZVnI1h1XMUw/edit?usp=sharing

Name: _____ Date: _____

YouTube Redesign Activity Guide

In this activity, it is the job of you and your group to redesign YouTube according to .

Step 1: Identify stakeholders. You will redesign Youtube with these stakeholders in mind.

Brainstorm <u>at least 10</u> stakeholders who care about YouTube's recommender algorithm.

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

Step 2: Identify some values of your stakeholders might want embedded in your algorithm. For each of your stakeholders above, identify at least two ways they care about the results of YouTube's recommendation system.

Stakeholder	Value #1	Value #1

Step 3: Fill out the following ethical matrix from your answers above:

Youtube (the company)				

Step 4: Given the ethical matrix above, identify the goal of your algorithm:

Why did you decide on this goal? Explain in a few sentences:

Step 5: Given the goal of your algorithm that you decided above, please describe what data you need to to teach your algorithm the correct actions:

Step 6: What features will your re-designed YouTube platform have to achieve this goal?

Step 7: Using the materials at your tablespace, design a mock up of the user's experience. Your prototype must show a mock up of:

- **Recommender system**
- **Comments section**
- **Search results**

Optionally, you can include mock-ups of:

- **Suggested search (auto-fill)**
- **Autoplay**
- **Advertisement settings**

YouTube Socratic Seminar

Description

Description	Learning Objectives	Time
Similar to AI Bingo, students in partners are tasked to recognize the various AI systems on the YouTube platform (e.g., advertisement matching algorithm, the recommender algorithm, comment classifier, etc) For each system, students identify what the algorithm is trying to predict and the dataset the algorithm uses.	1.d 2.b	30 min

Instructions

In a circle, give every student a copy of the article to read. Have students "popcorn read" the article.

After reading the article, ask students to summarize the article in their own words. A list of questions to discuss is provided below as a guide the conversation.

Reading

For this activity, students read an abridged version of the following article:

https://www.wsj.com/articles/youtube-under-fire-considers-major-changes-to-kids-content-11560953721?shareToken=st70732d59db684bc88a0e495597eb9f30

The article was edited to make it shorter and for the language to be more accessible to middle school students.

Socratic Seminar Questions

1. Can someone summarize this article?
2. What is the goal of this redesign? What is this platform optimizing for?
3. Can anyone name the stakeholders addressed in this article?
    a. Who is the most important stakeholder?
    b. Which stakeholder is making the most change or has the most power?
4. Do you think YouTube kids should be a separate product? Why? Why not?
5. Have you ever seen an inappropriate piece of content on YouTube? What did you do?
6. Would you use YouTube Kids app?
7. Do your parents make you use content controls now and
8. Would your parent like it?
9. Would a younger/older sibling like it?
10. How do you think advertisers feel about this?
11. Do you think it will be popular?
12. Do you think YouTube will lose profit? Is it okay if they lose profit?
13. What happens if there is more/less inappropriate content?
14. Why does autoplay exist? Who benefits from autoplay? Should autoplay exist?

Acknowledgements

Thanks goes to:

- Daniella DiPaola for developing the teacher guide for the YouTube Scavenger Hunt.
- Ilana Pelzman-Kern for building the AI Bingo cards.
- Cathy O'Neil for discussing her work on ethical matrices with us so we could use it as a teaching tool.
- Google's Creative Lab team for sharing Teachable Machines with us.

Support and resources to develop this work were provided by the MIT Media Lab Member Consortia and the Personal Robots Group directed by Cynthia Breazeal.

Students who piloted the curriculum pose for a "silly shot" outside of MIT's dome.