Machine Learning and Bias Activity

1. Go to [AI for Oceans](https://studio.code.org/s/oceans/stage/1/puzzle/1).   Recommend watching the video on the page before clicking the Continue button.
2. In this first time for the training set, make sure to choose all of the fish in the first 10 items (i.e. don't select any trash as fish).  Then click the continue button.
3. AI has been trained and will now be given items to identify (i.e. sort) as to whether they are fish or trash. Have it identify at least 20 items before stopping it.  Take a screenshot.  How did it do?

Graphical user interface

Description automatically generated

It identified several trash pieces as fish and identified fish as trash.

1. Click the Train More button (so you'll still be on Step #2 in the program) and do another 10 more items so AI has been trained on 20 items.  Then click the continue button.
2. Have it identify at least 40 items before stopping it.  Take a screenshot of the results.  Compare the 20 item training screenshot to how it did when it had only 10 trained items screenshot.  Did it improve?

A screenshot of a cell phone

Description automatically generated with medium confidence

It looks like it improved a lot more than the first time we trained it.

1. In Step #3 we're told the program is going to be used for a new purpose.  What is this new purpose?  Explain why we need to retrain AI for this purpose.

The new purpose is to identify whether fish belong in the sea or not. We would have to retrain it because the question and the data has changed. We would have to reprogram the AI to learn the new data sets like the starfish, crabs, octopus, etc. It will be able to identify more than just the regular fish that was in previous data sets.

1. Do the training for this purpose with 20 items.  Then click the continue button.
2. Have it identify at least 40 items before stopping it and clicking on the continue button.  Notice some items were incorrectly included and some items were incorrectly excluded.

* What are some of the items it incorrectly included (in the check results)?
  1. Coffee cups and soda cups
  2. Chicken legs
* Click on the NOT symbol (to the right of the checkmark) to see what items it excluded.  Which of these did it incorrectly exclude?
  1. Yellow fish
  2. No other item was excluded.

1. Click continue.  Based upon the video:

* What are 2 ways to collect training data?
  1. Asking users for training data
  2. Passively collect data from users
* What is the considered to be the code for machine learning algorithms?
  1. The data that is provided is the code for machine learning.

1. Click continue.  Pick one of the colors.  After 10 for the training data, click the continue button.
2. Allow it to test 40 items before stopping it.  Click the i icon in the top right corner.  Take a screenshot of the results along with the information provided by the i icon (i.e. how it determined what is included based upon the training data).  Click on a fish to see specifically how that fish measured up.

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated

1. Click continue.  Based upon the video:

* What are 2 ways that machine learning is used?
  1. Managing natural resources
  2. Creating medicine
* What are 2 risks for impacting society?
  1. Predicting crime, and false positives
  2. Evaluates applicants for college admissions, could deny people.
* How can we mitigate the risks?
  1. By using unprejudiced data/ unbiased data.
* Fill in the blank:  The most important aspect of machine learning is human learning.

1. Click continue.  Pick a word you want and state what it is.  Go through as many items of training data you think are necessary for AI to be able to correctly identify how you evaluate the word you chose.  Take a screenshot of the screen making sure to include the number you used in the training data.   Provide a justification of the number of training data used.

The word I picked was angry. I decided to use just 100 pictures to help identify if the fish was angry or not, because the more data sets, the more training the AI will get. Honestly, it requires way more than that. The data will not be good because it is based on my opinion of what fish are angry or not.

Graphical user interface, application

Description automatically generated

1. Click continue.  Provide a screenshot of what was included for your word and a screenshot of what was excluded from your word.  What characteristic(s) are there in the fish that were INCLUDED for your word?  What characteristics are there in the fish EXCLUDED for your word?

Graphical user interface

Description automatically generated

A screenshot of a cell phone

Description automatically generated with medium confidence

It seems that mainly the eyes, mouth and color were what determined what fish looked angry or not. If the eyes had downward slanted eyebrows and a frown it would mark it as angry. It looks like some of the happy or surprised fish were excluded, same as dorsal fins and tails. There are still some fish that was accepted as angry even though they don’t look like it.

1. Explain how this exercise is tied to the readings and related to your major/field. Remember to provide what your major/field is.

This exercise ties into the readings because it shows that we would have to train AI with data sets for it to be able to learn and identify things correctly. More importantly, this ties into the idea of the FANG vendors where their AI gathers information from the users by algorithms and behaviors that is collected from each user one-to-one. After noticing trends over time, the simple AI can recommend products or videos. This also ties into my field of Information Systems- Programming because this teaches us how when we code a machine to learn behaviors, it depends on the algorithms that we give it and other things like APIs and tools that the machine can use to help use the raw data that is given to it. The machine only can run as well as we program it to.