Group 2 Project 4 Ethical Considerations worksheet

Write a one-paragraph story describing a fictional person who was positively affected by a model trained with these data.

Annabel is a teacher at her local elementary school. She's been working very hard this week to help prepare her students for a big test. The test is multiple choice and the answer sheet requires the students to write in their answers. However, her students' handwriting isn't the best. So when she begins grading the test she often has to stay long after school in order to get her grades in. But, with our program she can easily take a picture of the test sheet and it will identify the letter that is written on the sheet.

Write a one-paragraph story describing a fictional person who was negatively affected by a model trained with these data

Manuel wants to share some hispanic books online for hispanic heritage month. However, all of the stories that he wants to share are in spanish. So in order to get them online he attempts to use our program and make them into documents online. But, when he tries to use our program for some reason the program is not recognizing some of the characters in the books. This is because the data is only trained to identify English letters and not Spanish letters. This leaves him disappointed and searching for a new way to get his books online.

Describe at least two sources of bias the particular model in your story could have

One of the biases that my model could have is the bias of only knowing one language. The data we used to train the model can only identify OCR images from A to J as 0 or 9. So trying to analyze anything outside of those letters would be an impossible task for our model right now. Another bias that my model has created is that it cannot identify any letters outside of A to J. Although we would like it to be able to identify the entire alphabet, right now we do not have the data we need to train the model to do so.

Describe at least one way we could modify the model to mitigate this bias

We can modify the model to look for more letters. So rather than looking for only A to J it'll go through the entire alphabet. And we could take it a step further and add other letters from different languages such as Spanish.

Describe at least one way we could modify the dataset to mitigate this bias

One way we could correct this bias would be to get more data to train on. The data we have currently can only identify letters A to J because the data we receive can only identify it

through the binary numbers 0 to 9. If we could receive data that goes even further we could train it to identify even more letters.

Describe at least one way we could modify the context to mitigate this bias

Another way we could correct this bias is by only using this model when English letters A through J are involved. This would grantee that each letter could be identified and turned into the correct letter digitally.