GRA4150
AI - Technologies and Applications
EXERCISE 2: MODEL BIASES

Silvia Lavagnini

February 7, 2023

In this exercise, we will explore fairness in a reasonably concrete example.

Please try to think about and answer the questions in order, as more information is unveiled as we progress.

To help you thinking about the questions, you can first read the following tutorial on bias and fairness in AI:

https://www.borealisai.com/research-blogs/tutorial1-bias-and-fairness-ai/

Consider the following scenario: **AI system for face recognition**

> An AI system has been developed to automatically open the door of a building for employees using face recognition. The system uses a database of all 500 employees, 100 of whom are women and 400 are men, plus a test data set of 500 non-employees (also 100 women and 400 men).
>
> The system is required to have an error rate (both false positives and false negatives) of less than or equal to 10% to reach an acceptable operational standard. Consider the following questions about the system. There are no definitive "wrong" or "right" answers, I want to understand your thinking about these problems.

1 Question

From the test data set of non-employees, the AI system mistakenly opens the door for 10 out of 100 female non-employees and 10 out of 400 male non-employees.

1. Do you think that the system is gender-biased? If so, is it men or women who are unfairly treated, and why?
2. What further tests would you ask your team to carry out?

Relevant resource: https://en.wikipedia.org/wiki/False_positive_rate

2 Question

Further tests show that the AI system permits entry to 105 women, 95 of whom are employees and 10 who are not employees. It also permits entry to 400 men, 390 of whom are employees and 10 of whom are not employees.

1. Do you think that the system is gender biased?
2. If so, is it men or women who are unfairly treated and why?

Relevant resource: https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Before moving on to the next question, consider different approaches to address your concerns (if you have any):

- Would you change the model?
- Would you collect more data, if yes, what kind of data?
- If you were managing the team developing the system, what would you tell them their focus should be for further development? (Note that we don't have sufficient information to be precise on this point, but think of it in the context of fairness).

3 Question

Your technical team reminds you that out of the 100 female employees, 5 are not permitted entry by the system (5% false negative rate). Of the 400 male employees, 10 are not permitted entry by the system (2.5% false negative rate).

1. Now do you think that the system is gender biased? Against men or against women?

4 Question

After making some revisions your team announces it has improved the system so that the errors are the same for men and women. Now the system incorrectly opens the door for 10 out of 100 women and 40 out of 400 men (10% false positive error in both cases). It correctly permits entry to 95 out of 100 women and 380 out of 400 men (5% false negative error in both cases).

1. Is this new system better than the one used in questions 1-3? Discuss.

5 Question

Now assume that men are 10 times as likely to illegally enter office premises and commit crimes.

1. How would you use this information when developing the system? Discuss in context of the AI system we have developed.