ENSF 611 Assignment 1

Aneesh Bulusu (UCID 30098046)

September 20, 2024

Resources used:

- https://www.prolific.com/resources/shocking-ai-bias
- Dr. Dawson's slides on "AI Ethics"

1. To me, AI Ethics means the practice of carefully auditing and reviewing the AI systems we use in our daily lives to ensure they are trained on reasonably unbiased data and are making predictions that comply with laws and social norms.
2. (a) I feel slightly uncomfortable after watching that video. As a person of color myself, I can imagine a scenario where such software is biased against me as well.
   (b) - Algorithmic bias can travel far and quickly: the facial recognition software that didn't recognize the speaker in the U.S. had the same issues as that in Hong Kong
   - You can teach a computer how to recognize other faces by showing it examples of faces and non-faces
   - Around 1 in 2 adults in the U.S. have their faces in facial recognition networks
3. (a) I have chosen example 3 ("US healthcare algorithm underestimated black patients' needs) at this link: https://www.prolific.com/resources/shocking-ai-bias
   (b) In this example, US hospitals used an algorithm that analyzed patients' healthcare cost history to predict which patients required extra medical care. Since it failed to account for the different ways in which white and black patients may pay for their healthcare, leading to incorrect and biased predictions which negatively impacted black patients.
   (c) This is an example of Selection bias, because the data that this model was trained on was likely that of white patients and failed to make distinctions between white and black patients, which existed regarding the different payment methods that were being used by the distinct groups of people.

(d) I picked this example as it seemed to be a very relevant extension of the example showed in the video, in particular regarding the ways that training AI models on data exclusively about white people can harm non-white people. Medical attention and paying hospital bills are a very relevant and controversial issue, and it should be a priority to use unbiased models if we are depending on AI systems to help us make judgements here.
(e) Include racial information, i.e., specify the race of the patient, in training data and when making predictions so the AI model is able to make more accurate predictions for black patients, and ensure that the proportion of black vs white patients in the training data accurately reflects the proportion of black vs white patients that are seen in hospitals.

4. (a) - Use ChatGPT to come up with idea suggestions for the course project ("Give me some examples of good machine learning projects..."
   - Ask ChatGPT a question about the pandas library because the documentation doesn't explain my issue as well as I would like ("How do I initialize an empty dataframe...")

   (b) I believe they should. AI systems are algorithms at the end of the day, and if companies are using material that doesn't belong to them to build their algorithms, which they then expect to profit from, then they should have a responsibility to compensate the individuals to whom that material belongs.

   (c) I have picked the "Job Search" issue from the "AI Ethics" slides. One way companies could address this issue is by scrubbing the gender from any historical data they are using to train their AI models, so the model does not make any distinctions between men and women. This way, the job search platform evaluates all candidates equally and does not propagate historical biases against women. I picked this issue as both the video from this assignment, as well as the example I chose in Question 3 are related to racial biases in AI systems, but it is important to also recognize that sexism is a problem we need to be aware of when providing training data to our models.