ASSIGNMENT 1

GERARDO GARCIA DE LEON
30172099

January 23rd, 2025

1. AI ethics to me means the conscious effort to ensure that AI is being developed, trained and used is in a ethical way that benefits majority of the population and society.

2. After watching the video I felt disappointed in the creators of the algorithm used for the facial recognition. My best guess is that this was developed and trained by an engineer or someone with a similar background. As engineers we are always attempting to find solutions to our problems that cover edge cases and help society as a whole. This video demonstrated that this algorithm was not trained to cover a variety of demographics, but rather one specific one.

   I learned that we can create datasets by choosing the data to train the model on. The bias that is shown on the model is a reflection of the bias in our choice of data selection. This can easily be solved if the data is more carefully selected and reviewed before using it.

   Another thing I learned was that the use of similar technologies travels around the world in a very quick amount of time. As the presenter mentioned, when she travelled to the other side of the globe and was demonstrated a similar piece of technology, I was as surprised as she was to find out they had used the same software, even on the other side of the world.

   Finally, I learned that this facial recognition software can be used in a variety of different fields to enhance the experience of the users. I believe the most useful use of this technology is for the police forces to be able to find criminals and recognize them with their software. This use would benefit the whole community and create a safer place for everyone.

3. https://www.bbc.com/news/technology-45809919
   https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

   In this example, Amazon attempted to train their model to help them filter out resumes and facilitate their hiring process. During the training, they used resumes from the past 10 years, most of which were male resumes. This lead the AI to believe that women's resumes were "less favourable" and created an algorithmic bias towards their resumes, impacting the likelihood of getting the job.

   This falls under the "Group Attribution Bias" as this model made assumptions based on the group that these people belonged to and attributing their group to be less desirable. There could also be a small bit of selection bias as the model was mainly trained on male resumes, but the distribution of male and female workers in amazon were about 60/40 in favour of males at this time.

   I picked this example because as a student who is applying to internships a lot, this has enlightened me to the fact that I may need to cater my resume to fit into an AI detection software before being reviewed by a human and changes my approach to applying.

This issue can be fixed by ensuring that the model is trained with an equal amount of male and female resumes and does not interpret certain demographics/groups as less desirable than others.

4. We could use generative AI to help correct any grammatical errors in our writing. AI can quickly sweep through the text and correct a misspelt word or an incorrect tense of a word. This should not be mistaken with text generation as that would be considered plagiarism. Another use could be to bring a different perspective on the material learnt in the course. Using generative AI to help summarize a topic after having reviewed it in the course can help refine the understanding of the topic by introducing a different explanation to the same material.

   I think they should pay for the material used as otherwise this would allow the usage of material to train the model to generate similar material and find a loophole around the copyright of the product as it's not the original. This gives the fair compensation to the creator of the product when any use is in play. If their product can be used to train an AI for similar creativity and generation without compensation, it disregards the future work of that artist/creator.

   The finance example where a loan approval system may unfairly judge applicants on different variables was briefly discussed in class. Companies can address this issue by taking some of the variables out that are less relevant to the problem at hand. The gender of an individual and their relationship status does not impact their finances as much as something such as spending habits and should therefore be considered less. I picked this issue because as an individual who has looked into taking out loans for purchasing a house, it is not easy to get approved as these types of systems may consider a lower score than if a human had manually done it as there's information that cannot be taught to a model such as spending habits previously mentioned. These can vastly differ between people of different types of economic, racial and marital backgrounds and heavily impact the ability of an individual to pay back a loan.

SOURCES:

[1] TED, "How I'm fighting bias in algorithms | Joy Buolamwini", *YouTube,* March 29 2017. [Online]. Available: https://www.youtube.com/watch?v=UG_X_7g63rY

[2] BBC, "Amazon scrapped 'sexist AI' tool", *BBC,* October 10, 2018. [Online]. Available: https://www.bbc.com/news/technology-45809919

[3] Jeffrey Dastin, "Insight-Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters,* October 10, 2018. [Online]. Available: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/