Robert Sanchez
Zhen Rui Ng
Bee Chang
Jonathan Schultz

Lab 13

1.

a) “Amazon scraps secret AI recruiting tool bias against women”

Exclude any gender related terms

- The model will disregard any gender related words during the training and selection process
- Review the datasets from the hiring process over a certain time period that may be male dominant to prevent any unintended bias

Hyperfixate Gender

- Set parameters to forcibly accept genders, race, and age at a certain socially valid ratio
- From a coding standpoint, model should incorporate fairness metrics into the algorithm to mitigate biases

Synthetic training data

- May have been trained with previous historic data that may have held unintentional biases
- Obtain truly objective data that is synthetic data to train the model

b) “Cleaning Up ChatGPT Takes Heavy Toll on Human Workers”

- Implement interactive blurring so that reducing moderators being exposed to harmful content. We could use combination of computer vision, sound and speech recognition, and text detection. Computer vision to identity harmful content such as: blood, violent behavior, and harmful activities. Sound and speech recognition to detect inappropriate language, unusual loud volume or sound made by a person or animal. Text detection to detect offensive language, attacks, or cyberbullying. These factors will be significant to classify whether the content has the higher probability of harming the psychological health of the moderator.

- Allow a language model to paraphrase the harmful content so the worker is not required to read the detailed harmful text in its entirety.

c) “Thousands of Authors Ask AI Chatbot Owners to Pay for Use of Their Work

- AI Chatbot owners should pay for use of the authors’ work if the authors are able to prove that their work were taken from piracy website. The use of the author work may or may not be the contributor to the lost of 40% of authors’ income, with IoT, online books are gaining popularity recent years, especially when COVID hit. So the lost of income may not caused by the AI chatbot owners using writing works to train their model.
- A system can be implemented to track and compensate authors whose work is being utilized by AI Chatbots. This might involve coding statements that enable a compensation mechanism based on usage metrics through royalties or licensing agreements.
- Overall limit the monetary gains of products developed by an AI. If the market desires an AI generated contents then ensure that they are properly labeled and accredited to original sources. Much like with music they must also purchase recording (model training) rights.

2.

Preventative measures

- Machine learning - From historical dataset or pilot report, what was the probability of survival before the pilot made a decision, what did the pilot do and what was the outcome.
- Deep learning - Decipher error codes, from recovered black boxes of crash sites, to recognize and detect in future flights. Such data as drops in pressure, engine failures, altitude loss, etc…
- Reinforcement Learning- Learn how to land the plane safely in different landscape or layout. Can help in optimizing better route planning and vehicle behavior based on real time conditions. Learn passenger survival probabilities for certain emergency landing maneuvers via simulation.
- Generative AI- Could aim in simulating diverse scenarios and testing purpose