

Ethics of Machine Learning & Artificial Intelligence

1. Complete the MIT short course in “Exploring Fairness in Machine Learning for International Development,” which may be found here:
<https://ocw.mit.edu/courses/res-ec-001-exploring-fairness-in-machine-learning-for-international-development-spring-2020/pages/module-one-introduction/>

NOTE1: In Module 4, do not concern yourself with the code in each of these examples. Instead, think about the methods presented conceptually to determine how they may be helpful for developing better models.

NOTE2: In the Module 4 gender example, they discuss hard voting and soft voting among multiple trained models. In this context, hard voting makes a prediction from multiple models by a simple majority vote for accuracy. Soft voting can only be done when all your classifiers can calculate probabilities for the different outcomes. In this scenario, soft voting makes a prediction from multiple models by averaging out the probabilities calculated by each individual model.

NOTE3: In the Module 4 natural language processing example, they discuss adversarial machine learning. Adversarial machine learning is a type of machine learning where two neural networks (type of machine learning algorithm) compete against each other. The first network tries to classify observations into one category, while the second network tries to fool it by creating fake examples.

2. After completing the course, answer the following questions:

Each answer will range from one to three paragraphs, but there is no maximum length.

- a. Can the framework developed by USAID be used in traditional organizations? Provide examples of each element in the framework in traditional organizational settings.
- b. Using the Solar Lighting example, discuss the pros and cons of using machine learning with historical data to determine credit worthiness.
- c. What is Fairness through unawareness? Discuss the pros and cons of this approach. What are the organizational risks for choosing “unawareness”?
- d. What is “demographic parity”? What is an example where you would want to use it?
- e. How do equalized odds and equalized opportunity differ? What is an example of where you would want to use one but not the other?
- f. What are some suggested techniques for mitigating gender bias?
- g. In the presented pulmonary health example, what is the key lesson learned about bias and fairness in trained models?
- h. In natural language processing models, what are some unintended causes of biases in models?

3. Read Chapter 7 in this book:
https://richmond.primo.exlibrisgroup.com/permalink/01URICH_INST/etnfu/alma9928471853506241
4. After completing the reading, answer the following questions:

Each answer will range from one to three paragraphs, but there is no maximum length.
 - i. What does the term “social mirror” mean and how does that impact fairness in machine learning?
 - j. How can an organization learn from Amazon’s failed algorithm? What can they do better? Should they scrap the idea of using machine learning and artificial intelligence for hiring? Justify your answer.
 - k. What is noisy data and how does that adversely impact making loan decisions?
 - l. What are the five ways that data mining discriminates?
 - m. In chapter 7.4, Hoffman mentions three tendencies. Identify and discuss those tendencies.
5. Using the knowledge gained from the MIT course and the readings, write a one-page executive summary for a CEO outlining how their organization can use machine learning ethically and fairly. This summary should be in the form of an executive summary with bullets and numbers and not a traditional essay.