**AIPI 510**

**Assignment 2**

This week we discussed data privacy and ethical issues that must be considered when working with data for modeling and some methods to mitigate them. In this week’s assignment we will put these skills into practice.

**Question 1 (30 points):**

A University decides to publish a book celebrating its “outstanding students” – those with GPAs in the top 10% in each major. The book has names and photos of students but does not contain address information or any other personal information. It does not contain their actual GPA, only that they were in the top 10%.

Does the book violate student privacy? Is it unethical in any way?

**Yes, it violates student privacy as, in the context, it did not mention the school had acquired permission from the students to release their personal information. Prior informed consent should be required.**

Would you interpret FERPA to ban this book? Do you think FERPA *should* ban this book?

**Yes. The privacy rule applies. FERPA should ban the book.**

**Question 2 (30 points):**

First, read the article [“Machine Bias”](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing). Then, answer the following two questions:

1. Explain how “learning bias” caused Northpointe’s software to have different levels of performance (and thus, caused disparate impact) in assigning risk scores to black defendants vs. white defendants, even though the software does not explicitly use race as a feature of its model.

**Learning bias occurred as some contextual information are not able to be included in the set of input values for the algorithm while the input values alone cannot represent the actual case, and thus fails to generate a rational output. Developing the algorithm to encompass more relevant features as well as feeding more similar cases to the training set would help reduce such bias.**

1. Explain how “deployment bias” came about in the case of states (such as Wisconsin) using Northpointe’s (and their competitors’) software?

**The author of the software stated that the intention of creating the software was to reduce crime rather than punishment. However, states like Wisconsin had used the software at every stage of the criminal justice process to sentencing convicts. Moreover, the software aggravated black convicts’ sentences regularly as it inaccurately predicted a convict’s re-offending rate. Improper application of the software created huge racial bias in the criminal justice system.**

**Question 3 (40 points):**

Find a news article about an incident where an AI product/system has been called into question due to potential ethical issues (similar to the case studies discussed in class). In **no more than** **two paragraphs**, briefly explain the following: (https://www.wired.com/story/twitters-photo-cropping-algorithm-favors-young-thin-females/)

1. Briefly summarize the incident and background

**Twitter’s photo-cropping algorithm favors thinner and younger-looking females.**

1. Explain why the ethical issue likely happened

**The racial and gender biases found reflects the biases of the human who contributed data used to train the model.**

1. Analyze whether the AI system meets each of the goals of 1) Fair, 2) Accountable and 3) Transparent AI

**Racial and gender biases appeared are a violation of fairness. Twitter has stopped using the algorithm right after its users’ report, hence accountability is considered. Some levels of transparency present as Twitter hosted a bounty challenge where the algorithm was provided for scrutinization.**