INFO 4390/5390 / CS 5382 Phase 1 Project Rubric

Spring 2024, Pierson & Koenecke

You will turn in your submission as a pdf. Please remember to tag ALL group members on your Gradescope submission. We expect to see the following items in each submission:

1. **Research Domain** (1-2 paragraphs)
   a. Specify both the broad application domain you're studying (e.g., "criminal justice" or "medicine") and the specific applications you're surveying in your literature review (e.g., "predicting recidivism risk scores" or "estimating knee pain")
   b. Specify the *sensitive feature(s)* you're focusing on (e.g., race, gender, or other demographic axes along which you're interested in exploring bias)
   c. Explain why this is an important research problem to study

2. **Literature Review** (~2 pages)
   a. For a team of size n (where $3 \leq n \leq 5$)*,* we expect you to cite at least n academic papers / investigative data journalism pieces in your research domain, using the ACM citation style. (So a three person group would need to cite at least three papers).
   b. For each paper cited/reviewed, please write a summary including:
      i. The category of paper
      ii. The contributions of the paper
      iii. How it fits in with the field's context
      iv. Comments on what you found good/bad about the paper (especially, e.g., the assumptions made in the paper)
   c. You should additionally explain how these n papers relate:
      i. To each other
      ii. To your research domain defined in Section 1

3. **Dataset Exploration** (~1 page)
   a. Report on what data source(s) you will use for your project to explore your research domain as defined in Section 1
      i. Include links / citations for your data source(s)
      ii. Explain where your data source comes from, and why you believe it is a trustworthy source
   b. Plan out what you will study with the data
      i. What specific research question will you be studying with these data?
      ii. What analyses do you plan to conduct to answer this research question?
         1. Is the dataset large enough to do these analyses? Does the dataset contain a variable for your sensitive feature of interest? *(If the answer to either of these questions is 'No', you should pick a different dataset.)*
      iii. Have other people used the data to do similar research? If so, what did they find? Do you expect to replicate results or find something different?
      iv. What are some limitations of your data?

4. **Contribution Notes** (1-2 paragraphs)
   a. For each team member's name, specify what their contributions were to this Phase 1 submission (e.g., who read which paper(s)).

5. **Bibliography** (as long as needed)

Note: it's okay if your final project (Phase 5) ends up on a different topic from what you submit for Phase 1; we understand that sometimes plans change. But, we still expect you to plan out in some detail what you intend to do for your final project as of this phase submission.