You are an ethics consultant hired to analyze the discrimination and fairness issues surrounding a machine learning classifier currently under development. Your job is to craft a policy memo that describes the classifier, discusses the fairness concerns that it raises, and offers at least one practical recommendation for mitigating those concerns.

Unlike in the previous policy memo, this time you are not expected to find a real-life case. I want you to *imagine* a context in which a machine learning classifier might be used to aid in decision-making. You should flesh out the case as much as possible: the more (imagined) detail, the better. Here are some examples:

- o A local school district is developing a classifier to predict which elementary-school students are likely to struggle in math. The district plans to preemptively enroll high-risk students in a supplementary math program.
- o A politician running for office is planning to distribute flyers about her candidacy. She hires a company to build a classifier that predicts which neighborhoods will be most receptive to the flyer campaign, so that she can concentrate her resources there.
- o FSU Basketball is building a classifier to predict collegiate athletic success on the basis of high school record. They'll use the classifier to decide which high school basketball players to recruit to FSU.

You may consider any plausible application that raises concerns about discrimination and fairness *except* for risk assessment tools used in the criminal justice system. We've talked a lot about COMPAS, so I want you to choose a *different* sort of case. You are allowed to select one of the other application we've talked about – e.g., loan approval, child abuse screening – but your analysis must go beyond what we have read for class in some way.

Your policy memo should contain the following sections:

1) **Background:** Describe your imagined scenario: What problem is the classifier being designed to solve? What decision(s) will it facilitate? Who will be using the classifier?

2) **Classifier Structure:** Sketch the basic structure of the classifier. Describe (a) the target variable, (b) the features the classifier will use to make the classification, and (c) the training data used to train the model.

3) **Fairness Considerations:** Use one or more of the texts we have read to unpack or analyze the fairness and discrimination issues at play in your imagined case study. For instance, you might describe how Barocas and Selbst's (2016) taxonomy applies to the case. You might explain the consequences of the fairness impossibility result for this particular application (Corbett-Davies et al. 2016, NIPS Tutorial). Or you might follow Hellman (2021) in raising a worry about compounding injustice.

4) **Practical Recommendations**: Formulate and defend one or more practical recommendations about how the fairness concerns you have raised in Section 3 should be navigated. For instance, you could propose that the model be trained on a particular dataset, or you could suggest that a particular sort of validation test be applied to the classifier once it is complete. Note that you need not take an overall

stand on whether the classifier should replace human judgment or not. You should shoot for quality over quantity in this section. If you only have space to thoroughly explain and argue for one recommendation, that is perfectly fine.

5) **References:** List all of the sources referenced in the policy memo. You may use any formatting style you like for your in-text citations and References section (APA, Chicago, etc.), as long as you are consistent. The References section should not be included in the word count.

Course Readings: Your policy memo should cite and make use of *at least two* of the readings that have been assigned in this class. Note that you need not agree with the texts that you cite.

Header: The header should contain the intended recipient of the memo and a brief description of the topic. The intended recipient of the memo is up to you: it could be the company building the classifier, the engineer working on it, or a decision-maker at the institution that will use it. The header does not count toward the word count.

Formatting and Length: Your policy memo must be 1,000-1,200 words (not including References and header). It should be written in 12-point standard font like Times New Roman, have 1" margins, and be double-spaced.

Deadline: Due on Canvas by 11:59pm on Sunday, March 27th.

Anonymous grading: Your policy memo will be graded by Marshall anonymously. Please do not put your name anywhere in the document that you upload. Canvas will link your submission with your name.

Plagiarism: Plagiarism of any kind (including mosaic plagiarism and patchwriting) will not be tolerated. Please see the university's Academic Honor Policy for more information. If you have any questions about what constitutes acceptable use of other people's work, do not hesitate to get in touch with me or Marshall.