4620 assignment 2

Summary:
Write a critical paper that revisits a recent controversy (~2000 words).

In a double-spaced paper, you will revisit a recent controversy involving different course themes. You can choose a case from a set of 5 predetermined cases. Your paper should draw extensively from the course materials, lectures, and in-class discussions, and

♥ present a comprehensive **plan** for undertaking the project in a way that <u>addresses ethical concerns and principles</u> such as *privacy*, *transparency*, *fairness*, or *value alignment*.

♥ The paper should also point to some **difficult tensions** and **unavoidable trade-offs**.
   ✡ You should explain <u>why it is not possible to satisfy</u> and
   ✡ have <u>all the ethical requirements</u> if the project is going to run.

♥ After that, you should provide **a thoughtful justification** for the <u>specific trade-offs</u> that you suggest.

You need to back up your claims with references from the readings in class or any other relevant scholarly work (peer-reviewed journal articles, conference proceedings, books). The end of your paper, therefore, should include a bibliography section (~5-10 references).

You can use the lecture slides to find about the scholarly, peer-reviewed works you want to cite. But presentation slides are not a source of scholarly, peer-reviewed work.

You can choose whatever style you like: https://www.bibliography.com/how-to/types-of-bibliography-styles/, but please stay consistent with your choice throughout the whole paper. All references need to be used as in-text citations. No bibliography=you would lose 5% of the total grade.

(1) HireVue has developed an AI product that allows employers to automatically assess job applicants who complete a video interview: https://www.hirevue.com/ The company promises to identify desirable candidates based on their diction, tone of voice, and speaking patterns as well as their facial movements and body language—a process that allows employers, according to the company, to pick up on cues overlooked by humans and reduce unconscious bias in the hiring process. The company has nevertheless been called out for perpetuating bias. Identify three reasons that could explain this critique. What might the company do, if anything, to address them? What challenges might they face in attempting to adopt these mitigations?

(2) In Loomis v. Wisconsin, a convicted criminal contested the use of the COMPAS risk assessment tool in his sentencing. He argued that the use of the tool violated his due process rights because trade secrets prevented him from interrogating the tool and because the tool suffered from bias. While the Wisconsin Supreme Court ruled against Loomis and the US Supreme Court refused to

hear the case, effectively condoning continued use of the tool, it remains extremely controversial: https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.htmlIdentify and explain three concerns that you have with the use of the tool, linking your normative objections to underlying technical problems. What techniques and tools might help address your concerns—and what dilemmas might these present? Recommend and justify a particular course of action.

(3) Charles Duhigg's 2012 account of Target's pregnancy prediction score has entered the canon of controversies in technology policy: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. Years later, however, there remains considerable debate about what, if anything, was objectionable about Target's behavior and what the company could or should have done differently. What are three main lessons that should be drawn from this story? In answering this question, identify what is the problem, both as a technical and normative matter. And drawing on this analysis, recommend how companies should behave in the future to address these concerns.

(4) Yilun Wang and Michal Kosinski released a paper that describes their attempt to use readily available AI methods to develop a model that could infer people's sexual orientation from photos of their faces: https://osf.io/zn79k/ The paper sparked considerable controversy for what it set out to do, for how it went about the research, and for how it presented its findings. In your own reading of the paper, what, if anything, is objectionable? Identify three things that the authors might have done differently. And what would be the best way to deal with their findings (i.e., with the possibility that others might develop and use such a model)?

(5) China's so-called "Social Credit System" has provoked both fascination and fear by outside commentators: https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion What are your three main concerns with such a system? What mitigations, if any, would help to address these concerns?