**Identify two different uses/applications of data in which biases in word embeddings might cause significant ethical harms, then briefly describe the specific harms that might be caused in each of the two applications, and who they might affect.**

The first harmful application I can think of is this tool being used for a potential person to hire for a job. If a person’s name is traditionally African American (Dion, Jamal, etc.), then it’ll be visually associated with words that the company will see as inappropriate for their “company culture”. Whereas with a typical white sounding name (Sam, Jack, etc.), might be seen as more attractive for a new hire. This would make it harder for people with more ethic sounding names land jobs they may be equally if not more than qualified for.

Another harmful application I can think of would be approving someone for a loan or something similar. The tool could scan a 25–35-year-old woman and match her up with words such as “stay at home mom”, “secretary”, or any other term that’s associated with a lower paid occupation. This preconception could prevent her from getting a loan, before actually applying for one, which would introduce a major sexism issue.

**Some researchers have designed ‘debiasing techniques’ to address the solution to the problem of biased word embeddings. (Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai, “Man is to computer programmer as woman is to homemaker? Debiasing word embeddings,” NIPS'16, 2016.) Such techniques quantify the biases, and then use algorithms to reduce or cancel out the biases that would otherwise appear and be amplified by the word embeddings. Can you think of any significant tradeoffs or risks of this solution?**

I believe the significant trade-offs would be that a lot of the main issue with certain biases (like pairing “black man” with “aggressive”) would be better mitigated. However, will this algorithm be able to quantify positive biases versus negative ones? If a certain type of person is associated with “strong”, “dependable”, or “intelligent”, would these biases also be reduced or cancelled out completely? This technique is a step in the right direction, but the risk of this solution is that its behavior seems way too vague.