EAST WEST UNIVERSITY

Mini Project - 3

Project title: Ethical Dilemma

Course Code: CSE487

Course Title: Cyber Security, Ethics and Law

Section: 01

Submitted to:

Rashedul Amin Tuhin
Senior Lecturer,
Department of Computer Science and Engineering

Submitted to:

Md Saiful (2019-2-60-040)
Fahad Ahmed (2020-2-60-174)
Kh. Foysal Ahmed (2019-2-60-003)

Date of Submission: 20 May, 2023

Artificial Intelligence Bias in Hiring Employee

Our Scenario: Suppose, Mr. Saiful works for a large technology company that is developing an AI-powered hiring system to streamline the recruitment process. The AI system uses machine learning algorithms to evaluate resumes and select candidates for interviews. The goal is to automate and **remove human biases from the hiring process**, ensuring fair and equal opportunities for all applicants.

However, during the testing phase, he discovers that the AI system is exhibiting bias against certain demographics. For example, it consistently ranks male candidates higher than female candidates for technical positions, even when their qualifications are identical. This bias could potentially lead to a significant gender imbalance in the company's workforce.

Ethical Dilemma:

The development team is under pressure to release the AI system to the market quickly, as it promises to revolutionize the hiring process. However, addressing and correcting the bias in the system will require additional time and resources but the office forced him to do the work. If he gives any objection then the company will suspend him from the job. The CEO is adamant about meeting the deadline, believing that the potential benefits outweigh the risks of bias.

Option 1: **Release the AI system**, he can choose to follow the CEO's directive and release the AI system despite knowing that it is biased. Doing so may cause perpetuating gender inequality and potential legal implications if the bias is discovered by external parties. However, he may face minimal backlash from his superiors and meet the deadline.

Option 2: **Delay the release** and address the bias: he can choose to voice his concerns to the development team and advocate for further testing and modifications to eliminate bias from the AI system. This may require additional resources and push back the release date, potentially upsetting the CEO and other stakeholders. But by doing this maybe it will suspend him from his job. However, by taking this approach, he can work towards ensuring a fair and unbiased hiring process, promoting equal opportunities for all applicants.

As an AI developer, he must decide which option aligns with his ethical principles and responsibilities. Will he prioritize meeting deadlines and potentially perpetuating bias, or will he prioritize fairness and equal opportunities, even if it means delaying the release and potentially facing opposition from company leadership?

Brainstorming phase:

People and organizations affected (Stakeholders):

- AI developer
- The CEO of the technology company
- The development teams
- The company's employees
- Job applicants (both male and female)
- Society as a whole
- The company Reputation
- Potential external parties, such as regulatory bodies or legal authorities

Risks, issues, problems and consequences:

- Perpetuation of gender inequality in the company's workforce
- Potential legal implications if the bias is discovered
- Damage to the company's reputation and public trust
- Negative impact on the morale and motivation of female employees
- Missed opportunities for qualified female candidates
- Decreased diversity and innovation within the company
- Negative societal impact, reinforcing stereotypes and biases
- Potential lawsuits or discrimination claims from affected job applicants

Who gets each benefit:

- Meeting the deadline and releasing the AI system on time
- Potential financial gains for the company
- Increased efficiency in the hiring process
- Potential reduction in human biases

Possible actions:

- Release the AI system as is, ignoring the bias concerns
- Delay the release and address the bias issues
- Conducted a quality assurance assessment
- Communicate about this concern with team, stakeholders, CEO
- Propose hiring part time developer
- Make a documentation of all events and serve as evidence of effort to ethical principles.

Analysis phase:

- **Responsibilities of the decision maker:** In our scenario, the decision maker is the AI tool hiring system. Actually, it's not the real decision maker, the ultimate decision maker is the programmer who made this system to function as autonomous. He has to make some decisions on the system, like when a candidate's CV will be accepted or rejected. To make these decisions, he has to consider some responsibilities by following general and professional ethics. The decision maker has a general ethical responsibility to act with integrity, honesty, fairness, avoid harm to individuals or groups, and give equal opportunities for all in his decision-making process. He should consider the broader societal impact of his decision and strive to promote the well-being of all stakeholders involved. He should design in a way that he is not doing harm to the society and human well-being. Now, as professionals in the field of AI development, the decision maker has specific ethical responsibilities outlined by professional codes of conduct and industry standards. These may

include principles such as ensuring transparency, avoiding bias and discrimination, respecting privacy, and promoting the responsible and ethical use of AI technologies.

- **Identify the rights of stakeholders:**

  In the given scenario, Job applicants have the right to equal opportunities in the hiring process and are evaluated fairly based on their skills, without facing any discrimination. Employees have the right to work in an environment that is free from discrimination, and harassment based on their gender, race, age or other characteristics. Company has the right to select and hire the most qualified candidates for the position based on fair evaluation. AI developers have the right to exercise their professional judgment and expertise in implementing AI systems.

- **Impact of the action options on the stakeholders:**

  First option is to release the AI system despite bias. This option will generate a negative impact for job applicants, particularly women. This system can harm the company's reputation and brand image. Developers who developed AI systems, their professional integrity and ethical standing may be questioned. Another option is to delay the release and address the bias. In this option job applicants have fair and equal opportunity in the hiring process. But the company will be affected.

Analyze consequences, risks, benefits, harms, and costs for each action considered:

1. Release the AI system as is:

The release of the AI system in its current state poses significant risks and consequences. It perpetuates gender inequality, potentially leading to a biased ranking of female candidates and negative impact on workforce diversity. Legal implications, reputational damage, and decreased trust in AI systems are likely outcomes. While it may meet deadlines and save costs, the harms include unfair treatment, systemic biases, gender imbalance, legal costs, and reputational damage.

2. Delay the release and address the bias:

Delaying the release to address bias offers benefits such as fair and unbiased hiring, equal opportunities, and a diverse workforce. It mitigates legal risks, preserves reputation, and upholds ethical principles. However, there may be stakeholder dissatisfaction, additional time and resources needed for testing and modifications, potential resistance from leadership, and increased costs.

After considering the consequences, risks, benefits, harms, and costs for each action, it becomes evident that delaying the release and addressing the bias (Option 2) yields the most favorable outcome.

- **Kant's, Mill's, and Rawls' approaches:**
  - In Kant's **"Categorical Imperative"** theory, he says that one should always respect the humanity of others, and that one should only act in accordance with rules that could hold for everyone. Kant believes in Absolute moral rule. Absolute moral rules mean the rules that everyone must follow in order to act morally. They are the rules that hold under any circumstances universally. Such as, we should never lie and we should never make any injustice. In our scenario, if we obey Kant's theory we will not be able to do any unethical thing. So the developer will lose his job and his family will be affected badly.
  - Mill's **"Utilitarian Theory"**, focuses on the results or consequences of our actions which produces the greatest good for the greatest number. There are two branches in Utilitarianism, one is act and the other is rule. In act we calculate utility just before the

event, on the other hand the rule is developed first based on utility with exception. It ensures equal opportunities for all job applicants and contributes to a more diverse and inclusive society.

- Rawls's **"Theory of Justice"** says, everyone in the society holds equal basic rights whether someone is from a minority group, but it does not matter. They also have equal rights and no one should deprive them. Rawl's "Theory of Justice" makes decisions based on "Veil of Ignorance" so all stakeholders are equal and there will be no discrepancy. Based on Rawls theory no chance for developer thinks about family, company or other things, all are equal.

- **Categorize each potential action:** The term "ethics" refers to a set of moral ideals. They have an impact on how a person makes choices and lives their lives. "Ethically obligatory" means that it is ethically essential to do the right thing and not to do the wrong thing. In our scenery, we have two options and none of them satisfy the ethical obligatory class. What is "ethically accepted" refers to what the majority approves. And the term "ethically prohibited" refers to activity that is expressly forbidden under an Ethics Code. In our dilemma, we have to delay building AI software.
  - Releasing the biased AI system as is: Ethically prohibited (violates general ethics of fairness and non-discrimination, and potentially professional ethics)
  - Delaying the release and addressing the bias: Ethically obligatory (upholds general ethics of fairness and justice, and aligns with professional ethics)

Decision Phase:

In the above dilemma, we have to choose between two options. Also, the Theory of Justice by Rawls is not practical for our case.

As per Kant's theory "rule is rule", if a developer does his duty, he will lose his job and his family will be affected so it will not benefit him. But the CEO is adamant about meeting the deadline, believing that the potential benefits outweigh the risks of bias.

Rawls' Delaying the release and addressing the bias supports the principles of justice as fairness, ensuring fairness and equal opportunities for all, especially disadvantaged groups.

Mill's approach: Delaying the release and addressing the bias promotes the greatest happiness for the greatest number, as it ensures equal opportunities for all job applicants and contributes to a more diverse and inclusive society.

So, we choose an option by following Mill's "Utilitarian theory". Because here the most ethical course of action is to delay the release of the AI system and address the bias. This approach prioritizes fairness, equal opportunities, and the elimination of bias from the system. Despite potential challenges and opposition from company leadership, upholding ethical principles is crucial in developing and deploying AI systems.