Software Engineering Ethics Algorithm Discriminates

Deep Mehta
College of Engineering
Computer Science
North Carolina State University
Raleigh, NC 27606
Email: dmmehta2@ncsu.edu

***Abstract*—This paper takes into consideration a fictional case study[1] of a software used by a company for screening the resumes of candidates and its disparate impact on minority communities and answers ethical questions of using such a system with the 7-step process[2] for ethical decision making.**

I. Problem Statement

Emporia, a large retailer had been experiencing high attrition rates in its sales department. Thus it decided to create a recommendation system that scores the candidates based on how likely are they to stay on the job longer.

This project was assigned to one of the developers within the company, Sandra. She applied Principal Component Analysis (PCA) to data from the resumes of current and past employees to identify the minimal set of features that best correlate with the length of tenure. Based on these attributes, the system was expected to list down the candidates as either 'Recommended' or 'Not-Recommended'.

One year after the system was implemented and used actively by the retailer, the following observations are made by Sandra (a) The company salesperson attrition rate falls by nearly 15 percent, (b) 92 percent of new sales hires are white.

Sandra is worried that because of the software recommendations, the company may be violating the legal standards for fair access to employment and thus decides to investigate this further. Sandra learns that the PCA listed zip code as one of the attributes that correlated with length of tenure. It recommended those candidates who lived in neighborhoods closest to the Emporia store, due to the reduced commute time they were more likely to stay on the job.

The problem is that these nearby neighborhoods consisted of mainly white, middle-aged people. The minority populations like Black and Latinos who constitute 80 percent of candidates applying for sales jobs, tend to live in areas that are further away from Emporia stores. This is the primary reason that the software is recommending these applicants as 'Not-Recommended'.

A. Sandra's Dilemma

Being a member of the Association for Information Science and Technology (ASIS&T), Sandra is well aware that it is her ethical and professional responsibility to make sure that the software she has programmed does not have a disparate impact on minority groups.

She shared her findings with Timothy who is the head of human resources. She mentions that the company may inadvertently be violating the disparate impact principle of the Civil Rights Act of 1964, which prohibits employers from using any employment practices that have unjustified adverse impacts on members of a protected class, such as lower-income persons, minority groups, or women. She recommends tweaking the software to use more appropriate metrics to recommend candidates.

B. Timothy's Response

Timothy is against the idea of changing the software in any way because it is working exactly as they anticipated it to, and has shown promising results. In addition to lower attrition rates, the company has also seen an increase in sales. Timothy isn't able to digest the fact that a computer software can discriminate, he believes that the application uses objective criteria to evaluate the best possible candidates for the company.

C. Broader Context

This problem may seem localized to the sales department at Emporia, but if we consider this from a broader perspective, there is a possibility that the success of this software will inspire the hiring team to start using it across all the departments. Furthermore, other companies might be curious to learn about the success and would eventually start using this or similar software as well. Thus, while considering the solution to this problem and its consequences, we should consider its global impact as well.

II. Facts

(a) In the past year, the attrition rate of the sales team has reduced by 15 percent.

(b) In the past year, 92 percent of new hires were white males/females.

While (a) & (b) seem correlated with the use of software, there is no definite way to say that these results are due to the software, there may be many other factors into play that year as well.

(c) Software considers zip code as one of the attributes while analyzing a candidate.

(d) Neighborhoods near emporium stores mainly constitute white, middle-aged people.

(e) Civil Rights Act of 1964 prohibits employers from using any employment practices that have unjustified adverse impacts on members of a protected class, such as lower-income persons, minority groups, or women.

(f) The developer of the software believes that the software is biased.

III. Stakeholders

The following stakeholders have been identified who will directly or indirectly be affected by the decision to use/change/recall the software:

(a) Sandra (Developer): Believes in ethical practices

(b) Timothy (HR): Wants what's best for the company

(c) Emporia (Company): Wants to increase profits and eventually its stock prices

(d) Employees at Emporia: Wants Emporia to stay in business so they can keep their jobs

(e) White population: Wants to get jobs in good companies like Emporia

(f) Minority population - Black, Latino, and other minorities: Wants to get jobs in good companies like Emporia

(f) Emporia's shareholders: Wants to see growth in the company and its stock prices

(g) Emporia's customers: Wants to buy good products at good prices

If you notice, none of the above stakeholders have any intentions of bias. They all have their own goals and expectations, but many times unknowingly one follows unethical practices to achieve those goals. As we see in this scenario, the use of this software is affecting the minority population but is beneficial to other stakeholders as it meets their expectations.

IV. Scenarios

A. Sandra reaches out to Timothy's manager and informs him/her of the current problem and why she is not happy with Timothy's response on this matter. The manager agrees and asks Sandra to update the software to eliminate the bias.

(i) Harm test: This option seems less harmful as the matter stays within the department. This may harm Timothy and even get him reported to the legal department.

(ii) Publicity test: Yes, this choice can be published in the paper since it seems like the right thing to do.

(iii) Defensibility test: Yes, this choice can be defended as it was the right thing to speak up and reach out to superiors.

(iv) Reversibility test: Yes, this choice will be an appropriate one if suitable steps were taken to eliminate bias and ensure such practices are not followed in the future.

(v) Virtue test: This showcases the virtue of honesty, courage, and intolerance.

(vi) Colleague test: Her colleagues would appreciate this choice to speak up, but would suggest she report it to the legal department as well.

(vii) Professional test: The professional committee would appreciate the decision of speaking up as soon as the issue was discovered.

(viii) Organization test: The legal department at Emporia would recommend raising an official complaint with them so appropriate actions could be taken to avoid any lawsuits.

Results: This choice is a safe bet if Timothy's manager agrees with Sandra and makes every effort to resolve the problem at the earliest. Sandra can go ahead and update the attributes to eliminate the bias and in addition to this spend enough time testing this software with people from various backgrounds, regions, and religions before releasing it to production.

B. Sandra reaches out to Emporia's legal department and takes their advice on this issue.

(i) Harm test: This option can be considered the least harmful of all the options. The legal department will take appropriate actions to fix the bias and ensure the company doesn't perform any unlawful activities.

(ii) Publicity test: Yes, this choice can be published in the paper since it's the ideal thing to do.

(iii) Defensibility test: Yes, this choice can be defended as it was the right thing to speak up and reach out to appropriate team.

(iv) Reversibility test: Yes, this choice will be an appropriate one if suitable steps were taken to eliminate bias and ensure such practices are not followed in the future. It would help if the incident was made public as a case study for other companies.

(v) Virtue test: This shows the virtue of honesty, courage and intolerance.

(vi) Colleague test: Her colleagues would appreciate this choice to speak up for the minority community and admitting her software was misconfigured.

(vii) Professional test: The professional committee would appreciate the decision of speaking up as soon as the issue was discovered.

(viii) Organization test: The legal department at Emporia would appreciate reporting the problem at the earliest.

Results: This choice seems like an ideal option to solve the problem at hand, as the unjust hiring practices will be eliminated and the minority communities will no longer be excluded from the hiring process. The legal department will also make sure the employees are educated about ethical biases to ensure such incidents don't occur in the future.

C. Sandra reports the unfair practices to U.S. Equal Employment Opportunity Commission[3]

(i) Harm test: This option seems the most ideal to get justice and stop the unfair practices followed at Emporia. This could harm Emporia with hefty fines, customer boycotts, reduction in share prices or even shutting down.

(ii) Publicity test: Yes, this choice can be published in a newspaper to set an example for other employees who notice

such unjust practices being followed in their companies but let it go unreported.

(iii) Defensibility test: Yes, this choice can be defended as Sandra has enough evidence to show that unjust practices were being followed using her software, and no efforts were made to fix it.

(iv) Reversibility test: Yes this seems like a good choice. It stops unjust practices, and punishing the company would set an example for other companies out there.

(v) Virtue test: This shows the virtue of honesty, courage, and intolerance.

(vi) Colleague test: Her colleagues wouldn't be in support of making it public as it might put the company at risk and in turn their jobs. They would instead recommend Sandra reach out to the legal department within the company.

(vii) Professional test: The professional committee would appreciate the decision of speaking up as soon as the issue was discovered, and report it to appropriate authorities when the company refused to rectify the errors in the system.

(viii) Organization test: The legal department at Emporia wouldn't be happy about it, and would have wanted Sandra to report to them first rather than directly going to the authorities, and may even sue her for violating the NDA.

Results: This option solves the problem at hand and gives justice to the minority communities who were the victims of unjust practices. This could disastrously harm Emporia and could even result in shutting down completely in the worst scenario which will in turn affect all the stakeholders, even the minority community because of the jobs being eliminated. Thus, even though this option might seem like a good one, it could harm every stakeholder along the way.

D. Sandra anonymously posts about these unfair practices on social media.

(i) Harm test: This choice may instantly tarnish the company's brand name and might even result in shutting down if customers start boycotting it, thus harming all the shareholders and employees. There is also a possibility that since it's anonymous, the post isn't taken seriously and no action is taken. This may harm Sandra if the company traces her to the post and sues her for violating the NDA (Non-Disclosure Agreement).

(ii) Publicity test: Yes, this choice can be published in the newspaper as the post is genuine and there's no intention of spreading hatred.

(iii) Defensibility test: Yes, this choice can be defended as Sandra has enough evidence to show that unjust practices were being followed using her software, and no efforts were made to fix it, and she was left with no other choice but to let the people know about these practices.

(iv) Reversibility test: Yes this would be a good choice if it had an impact on the company and its policies. If the post was ignored then it would be pointless.

(v) Virtue test: While this depicts the virtues of honesty, justice, and intolerance towards wrongdoings, it also depicts slight cowardice to post it anonymously. Understandably, Sandra might fear losing her job, but some situations might need her to speak up openly. Choosing such an option in extreme situations is fine, but shouldn't be chosen often.

(vi) Colleague test: Her colleagues may warn her about the risks of being caught and the chances of her getting sued by the company. Additionally, they might fear the loss of jobs since this option can tarnish the name of the company.

(vii) Professional test: The professional committee will respect the decision to speak up but they might not be pleased with the mode of delivery.

(viii) Organization test: The legal department at Emporia could sue Sandra for violating the NDA if they trace the post back to her.

Results: This isn't a foolproof solution, as pointed out earlier that the social media post might not be taken seriously since it's anonymous and no real action might be taken against the company.

E. Sandra updates the software in the background without anyone's knowledge and deploys it so that it works without any bias.

(i) Harm test: This option seems like a good option since it eliminates software bias without harming anyone directly. But, in the long term, it might affect Sandra's credibility if her superiors found out about her actions.

(ii) Publicity test: No, this choice shouldn't be printed in the newspaper as this is an unethical practice.

(iii) Defensibility test: It might be difficult to defend this in front of a committee as, even though her intentions were good, the implementation was unethical.

(iv) Reversibility test: Yes, this could be considered a good choice as this eliminated the bias in the current system, but this choice would fail to change the HR's mindset regarding bias and similar biases might occur in the future.

(v) Virtue test: While this shows virtues of courage and justice, this is also dishonest, and continuing to choose this option often wouldn't be a good idea and would affect Sandra's credibility with time.

(vi) Colleague test: Her colleagues would appreciate the ethical standpoint but wouldn't appreciate her updating the system through the backdoor.

(vii) Professional test: This would be considered unethical by the professional committee.

(viii) Organization test: The legal department at Emporia may take legal action against Sandra if HR raises a complaint against her.

Results: While this solution does solve the problem at hand, this isn't an ideal way. The HR has to realize the bias, take appropriate actions and involve the legal team letting them know of the unjust practices and prevent this from happening in the future.

F. Sandra doesn't think this is worth her time, as she is just a software developer and did what she was asked to do.

(i) Harm test: This option is harmful to all the involved stakeholders. In the short term, the minority population continues to be left out while considering employees for open positions. In the long term, the bias could be recognized by an external entity, which might result in hefty fines and a big blow to the company's brand name. This will harm Sandra's credibility as a software developer and she might have a hard time finding another job.

(ii) Publicity test: No, this choice is not suitable to be printed in the newspapers as it will affect Sandra's public persona, and she will be blamed for developing this biased software.

(iii) Defensibility test: No, this choice cannot be defended in front of the committee, since this was a result of software Sandra developed and in addition to that failed to raise concerns even after realizing the biases.

(iv) Reversibility test: No, this wouldn't be a good choice, as the minority population continues to be affected by these unethical practices.

(v) Virtue test: Sandra would become indifferent and selfish if she chose this option more often.

(vi) Colleague test: Her colleagues wouldn't be happy to know she didn't speak up even after she was well aware of the bias.

(vii) Professional test: The professional committee would consider this an unethical act.

(viii) Organization test: The legal department may hold Sandra responsible for the loss that the company would face if it is sued by an external entity, as she was the developer of the software.

Result: This solution isn't ideal and might never solve the problem. Even if it does, it can take a very long time to resolve the problem at hand. This solution will negatively impact most of the involved stakeholders in the short and long term.

V. The Final Choice

After considering option A-F in the previous section and evaluating the eight tests i.e. (i) Harm test, (ii) Publicity test, (iii) Defensibility test, (iv) Reversibility test, (v) Virtue test, (vi) Colleague test, (vii) Professional test, (viii) Organization test, for each of them, the most effective choice seems to be B. *Sandra reaches out to Emporia's legal department and takes their advice on this issue.*

The expected course of events is that the legal department officially files a report about unknowing unjust practices performed by the software and how they have rectified it. The legal department also issues mandatory training for all employees to spot unethical practices and report them.

A. Improvised software design

The current recommendation system should be updated to eliminate attributes like zip code, address, race, etc. i.e. all the attributes that directly or indirectly correlate with a particular set of populations and thus showcase bias in its output. The unbiased system would consider generic attributes like experience, previous manager's comments, etc.

VI. What's next?

Let's answer some questions in an attempt to avoid similar situations in the future

A. What could make it less likely you would have to make such a decision again?

As a software developer, one must follow the general principles for implementing autonomous and intelligent systems (A/IS)[4] as described by Institute for Electronics and Electrical Engineers (IEEE), to ensure the systems being developed are transparent, fair, inclusive, reliable, safe, private, secure, accountable and ethical.

B. What precautions can you take as an individual (announce policy on the question, change job, etc.)?

As an individual, one should set an example for others, so they learn from it and thereby avoid developing such erroneous software systems in the future.

C. What can you do to have more support next time (e.g., seek future allies on this issue)?

One must have a mentor or role model to look up to and reach out to for advice on such issues rather than trying to figure it out all on their own.

D. What can you do to change organization (e.g., suggest policy change at next dept. meeting)?

To start with, the following steps can be taken, firstly increase awareness within the organization about ethics and accountability while developing software. Secondly, involve a diverse set of populations in the development process of every software from the requirement analysis phase to ensure it is free of any bias against a particular group of people.

E. What can you do to change larger society (e.g. work for a new statute or EPA regulation)?

Such incidents must not be hidden away but taken up as case studies and be showcased in a list of bad examples and mistakes to avoid when developing software. I believe society learns from the mistakes of others. The above suggestion will help in avoiding unintentional biases. To put an end to intentional biases, strict policies must be in place to ensure every business or institution is compliant and follows the principles of Responsible AI[5].

References

[1] Online ethics center for education and science case study (https://onlineethics.org/cases/algorithm-discriminates)

[2] Seven Step Method for Ethical Decision Making. Michael Davis, 2013. (https://github.com/txt/se22/blob/main/etc/img/12steps.png)

[3] U.S. Equal Employment Opportunity Commission (https://www.eeoc.gov/how-file-charge-employment-discrimination)

[4] IEEE and Ethics (https://standards.ieee.org/wp-content/uploads/import/documents/other/ead/ead-for-business.pdf)

[5] Responsible AI principles by Microsoft (https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6)