I am a Machine Learning (ML) Engineer at a startup which is trying to speed up the hiring process for large organizations. Currently, large organizations get thousands of applicants everyday[1], but do not have enough staff and resources to review every one of them. We are developing a screening tool that takes resumes as input and decides whether an applicant's resume is strong enough to warrant a review by the hiring staff. More specifically, I am working on the ML algorithm that scores candidates.

As the algorithm I am working on determines the professional fate of thousands of people and the livelihood of their families, it is crucial that I focus on equality of opportunity to ensure my algorithm discriminates as little as possible based on sex, age, race, disability, and other "protected attributes" that could be inferred from a candidates' resume which do not determine their performance on the job.

ML algorithms work by finding existing relationships in data and using them for future predictions. Hence, I need to be aware that my algorithm will utilise the prejudice in our training data to make predictions during deployment. Amazon tried to create a similar system as ours and fed the algorithm resumes which were predominantly from men and found that their algorithm "preferred" male candidates by penalizing aspects on a person's resume that suggested they were women (for example, by penalizing the word "women's", as in "women's chess club capitan")[2]. As the engineer of the algorithm, it is my responsibility to understand whether our training data is representative of the demographic attributes of the candidates by developing tools to sort, group and graph our training data to understand patterns that are not obvious from individual data points. When I find areas where we fall short of representation in our training data, I need to work with relationship managers, who liaison with the companies we serve, to collect data that is needed to make our training data more representative. In this process, I need to be aware that other stakeholders (like the managers) might not understand how ML works. In such cases, it is my responsibility to educate them on the technicalities and consequences of the algorithm, and follow up to collect representative training data.

There is also a moral issue with algorithmic bias. As illustrated by the Amazon example above, although I can program my algorithm to disregard protected attributes, ML algorithms can see patterns in data that are difficult for humans to see and it is likely that the algorithm can indirectly infer protected attributes from keywords in candidates' resumes and unjustly dismiss candidates. From a moral standpoint, this is no different than the company having a blanket hiring policy to discriminate on the trait that my algorithm is discriminating on. While the Equality Act 2010[9] exists to prevent discrimination, a review by the Centre for Data Ethics and Innovation[10] has found "a regulatory environment with unclear requirements and weak enforcement" with regards to algorithmic decision-making and has advised the government to clarify "how existing legislation applies to algorithmic decision-making". Given I understand the ethical and moral consequences, I should be proactive and train myself and my colleagues on how to use Explainable AI tools like Google's What-If Tool[3] to understand how my algorithm makes decisions. When I find ways in which the algorithm discriminates, I should update my algorithm and work with relevant stakeholders to come to an appropriate resolution for the discrimination that might have taken place.

To train the ML algorithm, I am required to store personal data that is considered sensitive (data which reveals or concerns a person's protected attributes)[4]. In the process, there are blurred lines with respect to consent I could unintentionally cross. The principle of storage limitation in Article 5 of the General Data Protection Regulation[5] (GDPR) states that personal data must be stored "for no longer than is necessary for the purposes for which the personal data are processed". However, candidates' personal data will continue to be useful as training data so long as their data is reflective of the inputs processed by our system when it is deployed in the real world. Thus, if candidates are unaware that their data is used to train a ML algorithm, they may not expect their data to be kept for a prolonged period of time thus indirectly, but legally, infringing on their data privacy rights.

To mitigate this, I need to work with the companies we serve to ensure that opt-in consent forms clearly explain that gathered data will be used for the purpose of training a ML algorithm. However, the companies may prefer to hide this key piece of information in obscure locations in the consent form as they would not want people to perceive their brand as dehumanizing. Further, the management of my startup may also want the opt-out option to be hidden, as more data allows us to train better models which means we are able to offer better services which has a significant effect on our revenue. Navigating this situation is tricky and it is important to keep these alternate viewpoints in mind when convincing key stakeholders to do right by the candidates.

Lastly, storing data is a huge threat as data breaches are common (46% of 1,348 randomly surveyed businesses report having data breaches or attacks in 2020[5]). A breach could have a huge impact as resumes may include enough personally identifiable information (PII) to steal an applicant's identity which could have significant financial repercussions for them. To mitigate this, our system should anonymize the data by removing PII from candidates' resumes. This preventive measure falls in line with the "data minimisation" principle of the GDPR[6]. However, according to Article 7[7] and Article 16[8] of the GDPR, a person must be able to withdraw consent or update their data easily. Given these requirements, our system needs to encrypt data that is traceable to the person whom the data relates to and anonymize data where possible.

References

1. Biron, B., 2021. Amazon says it received more than 200,000 job applications for its 30,000 open positions. [online] Business Insider. Available at: <https://www.businessinsider.com/amazon-hiring-thousands-received-applications-2019-9?r=US&IR=T> [Accessed 22 October 2021].
2. Dastin, J., 2021. Amazon scraps secret AI recruiting tool that showed bias against women. [online] U.S. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> [Accessed 22 October 2021].

3. Pair-code.github.io. 2021. What-If Tool. [online] Available at: <https://pair-code.github.io/what-if-tool/> [Accessed 5 November 2021].
4. Information Commissioner's Office. 2021. Special category data. [online] Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/#scd1> [Accessed 22 October 2021].
5. GOV.UK. 2021. Cyber Security Breaches Survey 2020. [online] Available at: <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2020/cyber-security-breaches-survey-2020#fnref:1> [Accessed 22 October 2021].
6. Ico.org.uk. 2021. The principles. [online] Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/#:~:text=Article%205%20of%20the%20UK,lawfulness%2C%20fairness%20and%20transparency')%3B> [Accessed 22 October 2021].
7. Legislation.gov.uk. 2021. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (United Kingdom General Data Protection Regulation) (Text with EEA relevance). [online] Available at: <https://www.legislation.gov.uk/eur/2016/679/article/7> [Accessed 22 October 2021].
8. Ico.org.uk. 2021. Right to rectification. [online] Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/> [Accessed 22 October 2021].
9. GOV.UK. 2021. Equality Act 2010: guidance. [online] Available at: <https://www.gov.uk/guidance/equality-act-2010-guidance#overview> [Accessed 22 October 2021].
10. GOV.UK. 2021. Review into bias in algorithmic decision-making. [online] Available at: <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making> [Accessed 5 November 2021].