**Project Proposal Report**

Team name: LSJ

Team member: Jada Nguyen; Shirley Yao; Logan Hosoda

**What are you trying to do? Articulate your objectives using absolutely no jargon. Logan**

We’re aiming to examine and exploit the biases that currently exist in the automated resume filtering processes that are used at most companies to find qualified candidates. We’re attempting to propose a new model for the resume filtering process, create a small scale machine learning model to aid in recruitment, and ultimately improve applicant education on the current algorithms and practices in place.

**What is the problem? Why is it hard? Formulate your research questions. Are your research questions well formulated? They should be problem-driven and not method-driven. Jada**

In recent years, we've seen a surge in the use of AI algorithms in the hiring process, but AI tools can unintentionally discriminate against certain groups, reducing their job opportunities due to biased training data and flawed system design. This algorithmic bias not only causes unfair treatment, where job candidates are assessed based on biased criteria rather than skills and qualifications, but it also perpetuates systemic inequalities. As these biases are ingrained in hiring processes, they seep into organizational decision-making, reinforcing stereotypes and further marginalizing underrepresented groups, which can lead to biased products and outcomes.

It's difficult to address this issue because AI-driven hiring systems are complex and often rely on algorithms that are difficult to understand and identify correct sources of bias. This bias can be embedded in the training data, reflecting societal prejudices that require substantial effort to eliminate. Furthermore, the proprietary nature of many AI tools limits transparency, making external auditing and accountability challenging.

Research Questions:

* How does the diversity of training data used in AI-driven hiring tools impact their likelihood of discriminating against certain groups?
* Which aspects of AI-driven hiring tool design are most prone to introducing bias, and how does this affect job candidates from different backgrounds?
* What strategies can be used to reduce or eliminate bias in AI-driven hiring tools?

**Any preliminary evidence that you can answer the key research questions that you have in mind for your project? Any preliminary analysis/output? Jada**

Research from the US Equal Employment Opportunity Commission (Moore, 2023) has highlighted several ways that bias and discrimination can arise when employers use algorithms and AI in hiring and employment:

* **Overrepresentation in Negative Data:** Certain demographics are disproportionately represented in data sets that contain negative information like criminal records, evictions, and credit history. As a result, individuals from these groups face more significant barriers to employment and housing compared to their white counterparts.
* **Underrepresentation in Training Data:** AI algorithms trained on data lacking in diversity tend to be less accurate when evaluating underrepresented groups. This happens because the data doesn't capture the full range of experiences for minority groups. Without adequate representation, algorithms struggle to consider the unique circumstances that these groups face.
* **Proxies in Data and Inputs:** Even if algorithms don't explicitly include race, gender, or other protected categories, they often rely on proxy data that indirectly reflects these attributes. Examples of such proxies include zip code, names, schools attended, or online browsing history. These proxies can unintentionally introduce bias into AI-driven hiring decisions.

Although this highlights common sources of bias and offers potential solutions like debiasing techniques, it lacks specific examples, concrete data and empirical evidence. It also doesn't delve deeply into technical solutions.

**How is it done today, and what are the limits of current practice? Shirley**

The current practice to reduce algorithmic bias in the hiring process is the debiasing technique. One example is using this technique to remove personal information in resumes during the resume screening process (Clelland, 2023). In this example, the debiasing technique will remove personal information such as name, gender, race, and age, reducing direct bias based on demographic factors. However, this practice is limited in addressing indirect bias in other parts of the resume, such as educational background, work experience, or extracurricular activities. Consequently, hiring algorithms may still favor candidates from certain backgrounds or with certain experiences.

Another common current practice to reduce algorithmic bias in hiring is to include a diverse set of candidate data in the algorithmic training process (Clelland, 2023). If the algorithm undergoes training with a diverse set of resumes that involve candidates from diverse demographic backgrounds, it is less likely that the algorithm will exhibit bias against any specific group (Clelland, 2023). However, access to diverse training data may be limited, especially for underrepresented groups. This can result in imbalanced datasets that do not accurately represent the diversity of the candidate pool, which still leads to biases or inaccuracies in the training process.

**What’s new in your approach and why do you think it will be successful? Logan**

From our preliminary research, we found that there’s some bias within these algorithms and demographic data. Our approach aims to eliminate all demographic data from the equation, making our model evaluate a candidate purely based on skills and experience. The output will be an ordered list of most qualified to least qualified. There will also be an additional dictionary for retrieving candidate demographic and personal data for recruitment purposes. Recruiters will then traverse through the list of candidates, selecting the ones that they would like to interview while adhering to equal employment laws.

**Stakeholders/Who Cares? Jada**

Recruiters, applicants, job boards, professional profiles, and career advisers all play significant roles in AI-driven hiring processes. Recruiters are the decision-makers who rely on fair AI tools to help select candidates, while job seekers aim for equal opportunities without bias. Job boards shape job descriptions and affect applicant eligibility, potentially contributing to bias. Professional profiles, such as LinkedIn, are critical for networking and are often used by AI to evaluate candidates, making their accuracy key. Career advisers need to understand AI biases to guide their clients effectively. Meanwhile, students and entry-level candidates are often the most vulnerable to AI bias, emphasizing the importance of transparent and equitable hiring systems. Each stakeholder's needs and influence underscore the importance of addressing algorithmic bias in hiring processes.

**If you’re successful, what difference will it make? What impact will success have? How will it be measured?** **Shirley**

If our approach to reduce hiring algorithmic bias is implemented successfully, we will be able to help companies achieve a more fair and equitable hiring process. This ensures that all candidates regardless of their demographic backgrounds, have an equal chance for the job opportunity. Having candidates from diverse demographic backgrounds, companies will then have increased workforce diversity and inclusion, which can result in greater creativity. By ensuring fairness and inclusivity in the hiring process, organizations can improve employee satisfaction and retention rates (HiPeople, 2023). Reducing algorithmic bias can also help companies better comply with legal requirements related to anti-discrimination laws, such as Title VII the Civil Rights Act of 1964, which prohibits employment discrimination based on race, color, religion, sex, and national origin.

To measure if our approach is implementing successfully to reduce hiring algorithmic bias, we will use the following guidelines. To see if our bias reduction measures increase the hiring of diverse candidates, we will compare the selection rates of candidates from different demographic groups before and after implementing bias reduction measures. In addition, we will gather feedback from candidates about their experience in the hiring process, specifically their perception on fairness, transparency, and equal treatment. Another measure is tracking changes of the workforce diversity overtime to see the impacts of our bias reduction efforts. Lastly, we will assess whether our approach to reduce algorithmic bias helps companies better comply with legal requirements regarding equal opportunity employment.

**What are the risks and the payoffs? Shirley**

Although we attempt to train a machine learning model to reduce the hiring bias, there are some potential risks and payoffs that might limit our approach. First of all, the existing resume sample dataset we rely on to train our machine learning model may contain existing biases, such as the underrepresentation of minority groups. In addition, there is a risk that the machine learning model may overfit to the specific characteristics of the resume sample dataset, resulting in poor generalization to new and diverse applicant pools.

However, our approach offers several potential benefits. By eliminating existing proxies for demographic information and actively monitoring for biases, organizations can strive towards a fairer and more equitable hiring process. This not only helps in legal compliance with equal opportunity employment regulations but also fosters a more diverse and inclusive workplace culture. Additionally, implementing bias reduction measures in our machine learning model can enhance the efficiency and accuracy of the hiring process, leading to better candidate selection outcomes and improved organizational performance.

**How much will it cost (in terms of people, resources, etc.)? Logan**

With our proposed system, there will be a heavier emphasis on quality recruiters so that recruiters are able to evaluate candidates in a group setting to eliminate bias. Additionally, we will be attempting to make our machine learning model have as little memory as possible and will be storing all outputs and candidate data in the cloud. All of these are large scale costs that will be essential in the future but for the sake of this class’ timeline, the only necessary costs/resources are our group members, time spent on the project, and programming language used for the ML model.

**Plan of Activities - Plan of activities Provide a plan of activities and time estimates, per group member for the remaining quarter. Shirley**

The project will take place over the span of approximately 5 weeks with the expected plan of activities as follow:

|  | Logan | Jada | Shirley |
| --- | --- | --- | --- |
| Week 5 | 1. Refined project approach 2. Completed Project Pitch 3. CompletedProject Proposal | 1. Refined project approach 2. Completed Project Pitch 3. Completed Project Proposal | 1. Refined project approach 2. Completed Project Pitch 3. Completed Project Proposal |
| Week 6 | 1. Practicum Project Spotlight #1 2. Team discussion to finalize resume data set 3. Start to train ML model based on the dataset 4. Create dictionaries for the model | 1. Practicum Project Spotlight #1 2. Team discussion to finalize resume data set 3. Learn to create dictionaries for the model | 1. Practicum Project Spotlight #1 2. Team discussion to finalize resume data set 3. Learn to create dictionaries for the model |
| Week 7 | 1. Complete Project midpoint check presentation 2. Project midpoint report due on Friday - 10-May 3. Continue training the ML model | 1. Complete Project midpoint check presentation 2. Project midpoint report due on Friday - 10-May 3. Continue learning to train ML model and try best to help | 1. Complete Project midpoint check presentation 2. Project midpoint report due on Friday - 10-May 3. Continue learning to train ML model and try best to help |
| Week 8 | 1. Continue working on final project 2. Continue training the ML model | 1. Continue working on final project 2. Continue learning to train ML model and try best to help 3. Start to create a series of infographics that summarizes our project | 1. Continue working on final project 2. Continue learning to train ML model and try best to help 3. Start to create a series of infographics that summarizes our project |
| Week 9 | 1. Practicum Project Spotlight #2 2. Continue training the ML model | 1. Practicum Project Spotlight #2 2. Continue working on a series of infographics that summarizes our project | 1. Practicum Project Spotlight #2 2. Continue working on a series of infographics that summarizes our project |
| Week 10 | 1. Finalize the trained ML Model 2. Project final presentation 3. Final Project report Due on Friday | 1. Finalize the infographics 2. Project final presentation 3. Final Project report Due on Friday | 1. Finalize the infographics 2. Project final presentation 3. Final Project report Due on Friday |

**Mid-Term Exams Jada**

* **Key Milestones:** The completion of the midpoint presentation and report, this serves as a mid-term exam to assess the progress in training the ML model, ensuring that it's meeting the expected outcomes in terms of debiasing. This would include a review of key metrics, like diversity in the dataset and early model results
* **Testing and Validation:** Test the ML model to evaluate its performance in reducing bias. This could involve comparing the output of the model against known benchmarks or baseline tests.

**Final Exams**

* **Deliverables:** Trained ML model, infographics, formal presentation and report. This will encapsulate the entire project, demonstrating the research questions are addressed and achieving the goals of reducing algorithmic bias in AI-driven hiring tools.

**Literature Survey**

Raghavan, M., & Barocas, S. (2019, December 6). Challenges for mitigating bias in algorithmic hiring. Brookings. Retrieved from <https://www.brookings.edu/articles/challenges-for-mitigating-bias-in-algorithmic-hiring/>

The main idea of this paper is introducing algorithmic screening and how. It is useful for our project because it provides deep insight and examples into how the machine learning system “learns” to select candidates from previous datasets. In addition, it also discusses the effectiveness of algorithmic screening systems to reduce bias and its policy implications. Knowing this helps us gain understanding about the current practices and provide motivations for us to develop improved algorithms to reduce hiring algorithmic bias. However, similar to the “Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. Retrieved from”, it lacks detailed technical information and actionable guidance on implementing algorithmic de-biasing techniques, which could limit its utility for informing practical strategies to reduce bias in hiring algorithms. Our approach to overcoming these limitations involves conducting independent research to gather necessary technical details, enabling us to develop a machine learning model tailored to reducing hiring algorithmic bias.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. Retrieved from <https://dl.acm.org/doi/pdf/10.1145/3351095.3372828>

The main purpose of this paper is to provide an analysis into the bias related practices of algorithmic pre-employment assessments. There are some key research insights that provide academic evidence for our project. For instance, it mentions that algorithmic de-biasing techniques carry substantial implications for "alternative business practices" by automating the exploration for less discriminatory alternatives. This supports our end product as a machine learning model to reduce algorithmic bias in the hiring process as it can create positive implications to organizations, such as improved workforce diversity and inclusion. It also has useful information about the prerequisites for developing such de-biasing techniques, including transparency. This provides us guidelines when looking for a resume dataset we will use to train. There are several shortcomings of this paper. While this paper justifies the importance of de-biasing techniques, it does not provide specific steps or approaches about how to create or implement such techniques to reduce hiring algorithmic bias. In addition, while the paper offers valuable sociotechnical analysis and policy recommendations, it lacks detailed technical information and actionable guidance on implementing algorithmic de-biasing techniques. We aim to overcome these shortcomings by researching on our own to make up for the technical information so that we can still train a machine learning model that reduces hiring algorithmic bias.

Center for Democracy & Technology. (2020). Algorithm-driven hiring tools: Innovative recruitment or expedited disability discrimination? <https://cdt.org/wp-content/uploads/2020/12/Full-Text-Algorithm-driven-Hiring-Tools-Innovative-Recruitment-or-Expedited-Disability-Discrimination.pdf>

This paper explores the risks associated with the use of algorithm-driven hiring tools, particularly regarding disability discrimination. It discusses that these tools are designed to streamline recruitment, they can inadvertently discriminate against people with disabilities due to their reliance on certain criteria that may not accommodate a diverse range of abilities. It brings to light a specific type of discrimination that may not be widely recognized, providing an additional angle from which to examine algorithmic bias. But while the paper focuses more on identifying discrimination it doesn’t deeply explore technical solutions to mitigate it.

Frissen, R., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach recognize bias and discrimination in job advertisements. *AI & SOCIETY*, *38*(2), 1025–1038.<https://doi.org/10.1007/s00146-022-01574-0>

This study was conducted to test how resume filtering machine learning models discriminate against certain applicants or even dissuade ceratin demographics from applying to a job listing. The study categorized harmful discriminatory language into the categories masculine-coded, feminine-coded, exclusive, LGBTQ-coded, and racial language. We think that this article will not be useful to the development of our model specifically, but is beneficial to have knowledge about regardless, thus solidifying our concept of eliminating all personal biases from the filtering process. It also ensures that when we do look at job descriptions, to observe what biases are written into them and how it will affect the output of our model’s data. We like the categorical nature of the study and that it will be beneficial to implement similar categories when coming up with the criteria for our sentiment analysis.

Dobson, S. (2018). Is Ai Biased in Recruitment? *Canadian HR Reporter*, *31*(12), 1,8.

This short podcast highlights the challenges associated with AI/ML algorithms and tools in the recruitment process. This particular issue arose after Amazon’s recruitment tools were found to contain biases against marginalized groups such as women. The recent development has many questioning the true fairness of the recruitment process. This article will primarily be used for supplementary sources and further strengthening our argument that these tools should be kept as is and not a complete replacement of recruiters altogether. They also push the narrative and importance of having data that is representative of the parties involved, which is also something that we’re trying to implement into our model.

**References**

Center for Democracy & Technology. (2020). *Algorithm-driven hiring tools: Innovative recruitment or expedited disability discrimination?*<https://cdt.org/wp-content/uploads/2020/12/Full-Text-Algorithm-driven-Hiring-Tools-Innovative-Recruitment-or-Expedited-Disability-Discrimination.pdf>

Clelland, K. (2023). Algorithms in Hiring: Addressing Bias in Candidate Selection. Career Insights. Retrieved from <https://www.linkedin.com/pulse/fairness-hiring-mitigating-bias-recruiting-algorithms-kris-clelland/>

Dobson, S. (2018). Is Ai Biased in Recruitment? *Canadian HR Reporter*, *31*(12), 1,8.

Frissen, R., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach to

recognize bias and discrimination in job advertisements. *AI & SOCIETY*, *38*(2),

1025–1038.<https://doi.org/10.1007/s00146-022-01574-0>

Fuller, J. B., Raman, M., Sage-Gavin, E., & Hines, K. (2021). *Hidden workers: Untapped talent*. Harvard Business School.<https://www.hbs.edu/managing-the-future-of-work/Documents/research/hiddenworkers09032021.pdf>

HiPeople. (2023). Improved Employee Satisfaction: Employees in Diverse Work Environments Experience Higher Engagement and Commitment Among Employees. HiPeople. Retrieved from <https://www.hipeople.io/glossary/diversity-hiring>

Moore, C. (2023, January 31). Navigating employment discrimination with AI and automated systems: New challenges and solutions. In *Meeting of the U.S. Equal Employment Opportunity Commission*.<https://www.eeoc.gov/meetings/meeting-january-31-2023-navigating-employment-discrimination-ai-and-automated-systems-new/moore>