
Identifying Solutions to Discrimination in Artificial Intelligence

Jared S Chapman

Northwestern University

www.linkedin.com/in/jared-chapman-msds

<https://github.com/jschapym>

jared.chapman100@gmail.com

ABSTRACT

Artificial Intelligence (AI) has the potential to improve efficiency, reduce bias, and promote inclusivity in resume screening. However, there are also concerns regarding the unintentional perpetuation of existing societal biases. This paper investigates the ethical and technological challenges of AI-driven resume screening. Specifically, it investigates how historical data—reflective of previous hiring practices—can unintentionally instill biases into AI systems. This can inadvertently lead to discriminatory outcomes. The increasing use of AI in recruitment processes has revolutionized the way organizations assess candidates. Specifically, in applicant tracking systems and resume screening. AI systems utilize Natural Language Processing (NLP) and machine learning algorithms to streamline the hiring process. These systems do so by analyzing large volumes of resumes to identify the best candidates for a specific position based on objective criteria. This research utilizes techniques such as Named Entity Recognition (NER) and gender and ethnicity prediction to analyze potential biases in AI recruitment models. The findings contribute to a deeper understanding of the ethical implications of AI in recruitment. Recommendations for creating AI systems that are efficient and equitable will ensure fair opportunities for all candidates.

AI-Driven Resume Screening Steps

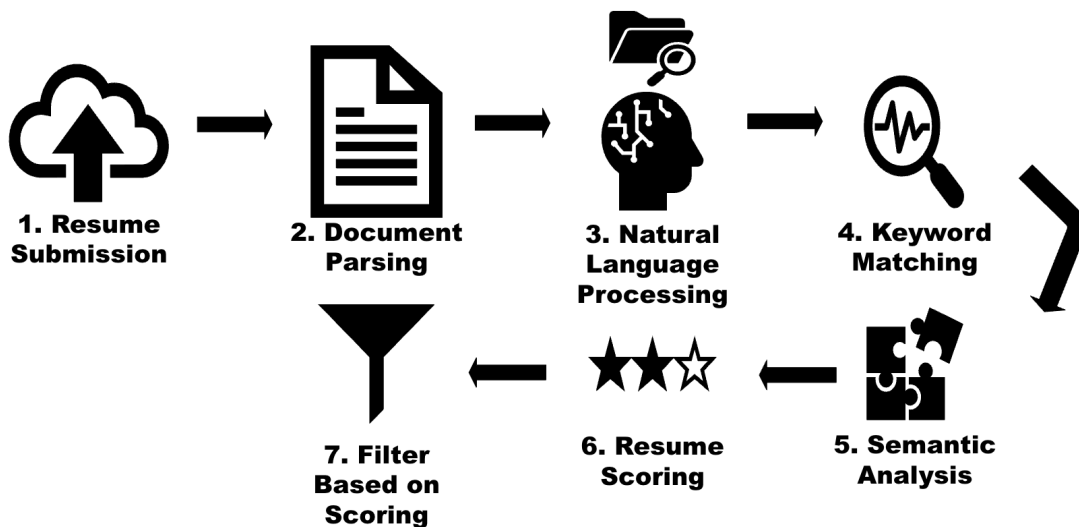


Table of Contents

ABSTRACT	i
INTRODUCTION AND PROBLEM STATEMENT	1
BACKGROUND	3
AI in Resume Screening	3
Bias in AI Systems and Mitigation Thereof	4
Ethical Implications of Biases	6
DATA	7
Figure 1.	8
Figure 2.	9
METHODS	10
Data Preprocessing.....	10
Name Extraction	10
Gender Prediction	11
Ethnicity Prediction	11
K-Nearest Neighbors Classification	11
RESULTS	12
Figure 3.	12
Figure 4.	12
Figure 5.	13
Figure 6.	14
Figure 7.	15
DISCUSSION	15
K-Nearest Neighbors Analysis	15
Gender and Ethnic Language.....	16
Avoiding Discrimination	16
CONCLUSIONS	17
DIRECTIONS FOR FUTURE WORK	17

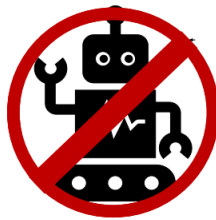
ACKNOWLEDGEMENTS	18
DATA AVAILABILITY	18
CODE AVAILABILITY	18
REFERENCES.....	20

INTRODUCTION AND PROBLEM STATEMENT

Bias is a significant concern in AI-driven resume screening and recruitment processes. A prime example of this issue was documented in the American multinational technology company: Amazon. Amazon introduced an AI recruiting tool that was abandoned in 2017. This was due to an issue in unintentional discrimination.

The tool was trained on resumes that were submitted to Amazon over a 10-year period, which were predominantly male applicants applying for technical roles. This resulted in the AI system inadvertently prioritizing male-dominated language and experience. Many male engineers used words such as “executed” or “captured” more frequently than women. This penalized female applicants based on the language and wording they used more frequently when constructing their resumes (Dastin 2018). Even though there were not explicit biases in mind when training this recruiting tool, the training data was unintentionally biased.

How Amazon's AI Recruiting Tool Perpetuated Bias



Amazon's AI recruiting tool was abandoned in 2017 due to unintentional bias against women.



The AI was trained on 10 years of resumes from mostly male applicants for technical roles.



The system was influenced by language patterns in male resumes. Specifically, words like 'executed' and 'captured.'

Female applicants were unfairly penalized for using collaborative language, while male resumes were favored.



The tool was not intentionally biased, but the data reflected historical hiring trends, leading to adverse effects.

To avoid biases, AI systems should be trained on diverse data, regularly audited, and involve human oversight.

The increase in adoption of Artificial Intelligence (AI) across numerous industries has introduced transformative changes, particularly in the human resources field. One significant application of AI in this field is in Applicant Tracking Systems (ATS) and the resume screening process. AI-driven systems are more prominently being instrumented to filter and evaluate a candidate's qualifications. This can be extremely efficient in terms of streamlining the hiring of new employees, while eliminating human error.

AI can analyze vast amounts of hiring data (i.e. resumes) and provide data-driven recommendations on candidates that meet the position requirements. Using Natural Language Processing (NLP) and machine learning algorithms, these systems can identify patterns and make predictions based on candidate qualifications and experience.

AI resume screening presents numerous advantages for organizations. They can reduce the time and cost associated with reviewing hundreds to thousands of applications by swiftly analyzing resume and candidate data in a fraction of the time that a human could. Theoretically, this would also decrease the bias and judgement coming from humans, such as gender, age, or race (Binns 2018). AI focuses on objective criteria, such as educational background, years of experience, and keywords associated with the job description. This gives potential to reducing the discrimination that may come from human judgement and to make the recruitment process fairer and more inclusive.

However, even though AI proves to be more efficient and to provide more objectivity, it still comes with some challenges. The algorithms used in AI screening learn from historical data, which is typically derived from previous hiring decisions. If the data used to train these AI systems is biased, specifically, if the data reflects existing inequalities in the hiring process, they may unintentionally perpetuate these biases.

For example, if the historical data reflects a male-dominated workforce, the AI may favor resumes that reflect male-oriented qualifications, which, in turn, may disqualify female candidates. Data-driven discrimination raises ethical concerns, especially since there is a rise in instrumentation of AI to make decisions in the hiring process. This ultimately can affect the careers and livelihood of many individuals.

The goal of this research is to investigate the potential for discrimination in AI resume screening and applicant tracking systems. Specifically, exploring how biases in training data can result in discriminatory hiring practices, even when the AI systems are not designed to be biased. Examining the ways in which AI can perpetuate existing biases will direct this study to identify strategies to promote best practices. This can assist in the mitigation of discrimination in AI-driven hiring processes.

This research will contribute to a more thorough understanding of the ethical implications of AI in recruitment and propose potential solutions to ensure fairness and inclusivity in these processes. With AI becoming an integral part of the recruitment process, it is necessary for organizations to understand the potential risks associated with these technologies. This research will highlight ethical and technological challenges of AI-driven resume screening, ultimately striving to provide recommendations for creating AI systems that are efficient and equitable.

BACKGROUND

The overlap of AI and human resources is a quickly evolving field with several issues, specifically in bias and discrimination. The following review will focus on literature pertaining to the AI recruitment process, specifically discussing a few main themes:

- The benefits and challenges of AI in resume screening
- The mechanisms in which bias can enter AI systems
- The ethical implications of these biases

AI in Resume Screening

AI has the potential to revolutionize recruitment practices by automating time-consuming tasks, improving efficiency, and enabling data-driven decision-making. The instrumentation of AI in recruitment, specifically resume screening, has become extremely common in modern society. Automating the initial stages of the hiring process allows thousands of resumes to be sorted through and examined in a few minutes. This then identifies top candidates based on predefined criteria in data, such as education, experience, skills, and qualifications (Hao 2020). This approach is a method to save time and resources, especially for companies that receive high volumes of applications.

A significant advantage of AI resume screening is the reduction of human error and unconscious bias thereof. Human recruiters may unintentionally favor candidates who share similar characteristics, backgrounds, and experiences with them. They may also unintentionally disfavor individuals who fit into certain stereotypes (Green 2019). AI should, in theory, mitigate these issues by focusing on objective and data-driven criteria. Studies have suggested that AI could level out the playing field for candidates by removing subjective judgements based on gender, race, appearance, or age. This would ensure that candidates are assessed solely on qualifications and aptness for the position (Binns 2018).

Bias in AI Systems and Mitigation Thereof

While AI has a great potential for objectivity, it still relies on historical data for training, which introduces flaws of data bias. Machine learning models used in AI resume screening are trained on large datasets, often reflecting existing societal biases. These biases in data can unintentionally cause the AI systems to rate candidates differently than they otherwise would. This can result in AI algorithms favoring resumes that reflect patterns of historical hiring practices, which inherently perpetuates discrimination against certain underrepresented groups (Angwin et. al. 2016). This can lead to the exclusion of qualified candidates who may not necessarily fit into the patterns of the training data.

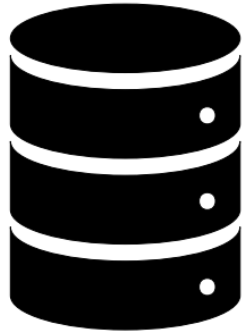
One main issue with these AI systems is that it is difficult to distinguish the logic of the decision-making of them. A “black box” is an AI system that is not transparent to users, making it difficult to trace the exact logic behind a certain decision (O’Neil 2016). If the transparency is improved, it may increase the accountability and understanding of these systems, making it easier to determine how to better algorithms to decrease bias.

There have been several strategies introduced to mitigate biases as the awareness of them has increased. Using diverse and representative datasets to train AI models is one main solution to these concerns. By confirming and ensuring that the datasets used to train these systems contain examples of a wide variety of candidates, the likelihood of an AI system favoring one candidate over another could decrease immensely (Mehrabi et. al. 2019). The solution to this could be simply ensuring that the resumes in a training dataset represent many experiences, education levels, and career paths.

This also requires an algorithm to be more flexible with terminology. This, alongside several other algorithmic fairness techniques, have been instrumented to mitigate the biases in machine learning models. For example, fairness constraints are mechanisms that ensure fairness in a wide variety of classifiers in a principled manner. These are used to decrease the disproportionality of the AI systems, preventing disadvantage to any demographic.

Similarly, adversarial debiasing is a machine learning technique that works to reduce bias in algorithms by simply training two neural networks that are completely opposite of one another. These include the classifier network, which is trained to perform the primary task, and the adversary network, which trains to identify biases. For example, the classifier may identify if a candidate is qualified while the adversary may identify whether the candidate is of a certain race (Zafar et. al. 2017). These techniques offer interesting and effective mitigation strategies to the unintentional biases of AI systems.

Database Bias



Data used to train an AI model is biased because it is underrepresenting or historically skewed.

i.e. A hiring algorithm trained on resumes from men may not identify equally qualified women.

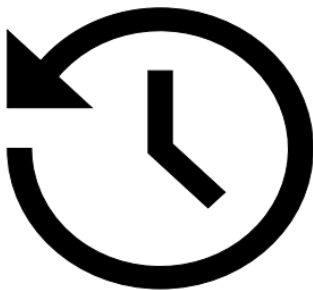
Labeling Bias



Data used to train an AI model is incorrectly or inconsistently labeled, leading to flawed predictions.

i.e. If a training dataset includes mislabeled names of individuals from different ethnic groups, the model may misidentify them.

Historical Bias



Historical biases present in the data can perpetuate discriminatory outcomes in AI predictions.

i.e. An algorithm trained on a historical dataset of predominantly white men, the language may be different from those of other demographics. This may result in unintentional discrimination.

Confirmation Bias



The system unintentionally favors resumes that align with pre-existing assumptions or patterns.

i.e. If an AI prioritizes certain keywords, such as "leadership," it may overrate resumes that emphasize those keywords and disregard candidates who do not use the same phrasing.

Ethical Implications of Biases

One of the key issues concerning ethics in AI resume screening is the lack of human empathy and context in these systems. AI algorithms do excel in processing giant volumes of data. However, this does not give them the ability to understand personal context, a candidate's potential outside of their resumes, or cultural fit. Since these characteristics and attributes cannot necessarily be measured or quantified, they typically are not captured in the screening processes. Yet, these measures are critical to understanding the success of candidates in a company's culture. With the reliance on these systems increasing, a more mechanical approach, being inconsiderate of the complexity and diversity of a human's talent, could become the new norm.

The use of AI in recruitment stages raises concerns regarding transparency and accountability. It is essential to ensure that the systems in place for candidate tracking and hiring are transparent, explainable, and able to grow. For instance, a candidate should have the access to information surrounding the reasons behind why they were either selected or rejected. Organizations should also hold accountability for the potential of discrimination that comes with these AI systems.

The EU's General Data Protection Regulation (GDPR) is an example of a framework that is beginning to address the transparency in automated decision-making. Specifically, Article 22 deals directly with decisions made by automated processing. It is stated that individuals should not be subjected to decisions based solely on automated processes, except in specific circumstances. These include the following: The decisions must be necessary for the performance of a contract, the decisions must also be authorized by union or member state law, and an individual must give explicit consent to the processing. It also stipulates that the companies must provide information surrounding when these processes are being instrumented (Gillespie 2018).

In the United States, local jurisdiction have started to address the ethical concerns with AI in recruitment processes. New York City's Automated Employment Decision Tool Law was mandated in 2023. This law requires employers who use AI driven tools to audit these systems annually for bias, ensuring they do not contribute to unfair discrimination of candidates. This requires the companies to audit these tools with third-party systems as well (New York City Department of Consumer and Worker Protection 2023).

Similarly, California has introduced several privacy laws in the past decade, focusing more on data privacy than AI systems in hiring processes. The California Consumer Privacy Act allows for residents to opt out of data collection practices, including those of AI-automated processes. The California Privacy Rights Act introduced separate rights that allows residents to access and delete their personal information. This also grants them the right to transparency about automated decision-making processes that may affect them. These laws show commitment to the

balance of innovation with AI systems and fairness in them (California Legislative Information 2020). With the growth in new laws addressing the ethical implications of AI, an improvement in these systems make come with time.

DATA

The dataset used in this study contains 13,389 datapoints, each representing an individual resume. The dataset was sourced from the Resume Dataset available on Hugging Face (InferencePrince555). This contains resumes categorized by job titles, including the text of the resumes. Each datapoint consists of two attributes: resume category and resume text. The resume category refers to the job title or role (i.e. “accountant”) and the resume text includes detailed content of the resume. The text includes a candidate’s name, work experience, education, skills, etc.

Prior to analysis, preprocessing of the data was completed to clean the set. This included removing special characters, multiple spaces, and other formatting issues. Any missing entries were excluded to ensure a clean and reliable dataset for analysis. One of the main features of this dataset is the distribution of job categories. A table outlining each of the categories and a count for the number of resumes included in each category can be found in Appendix A.

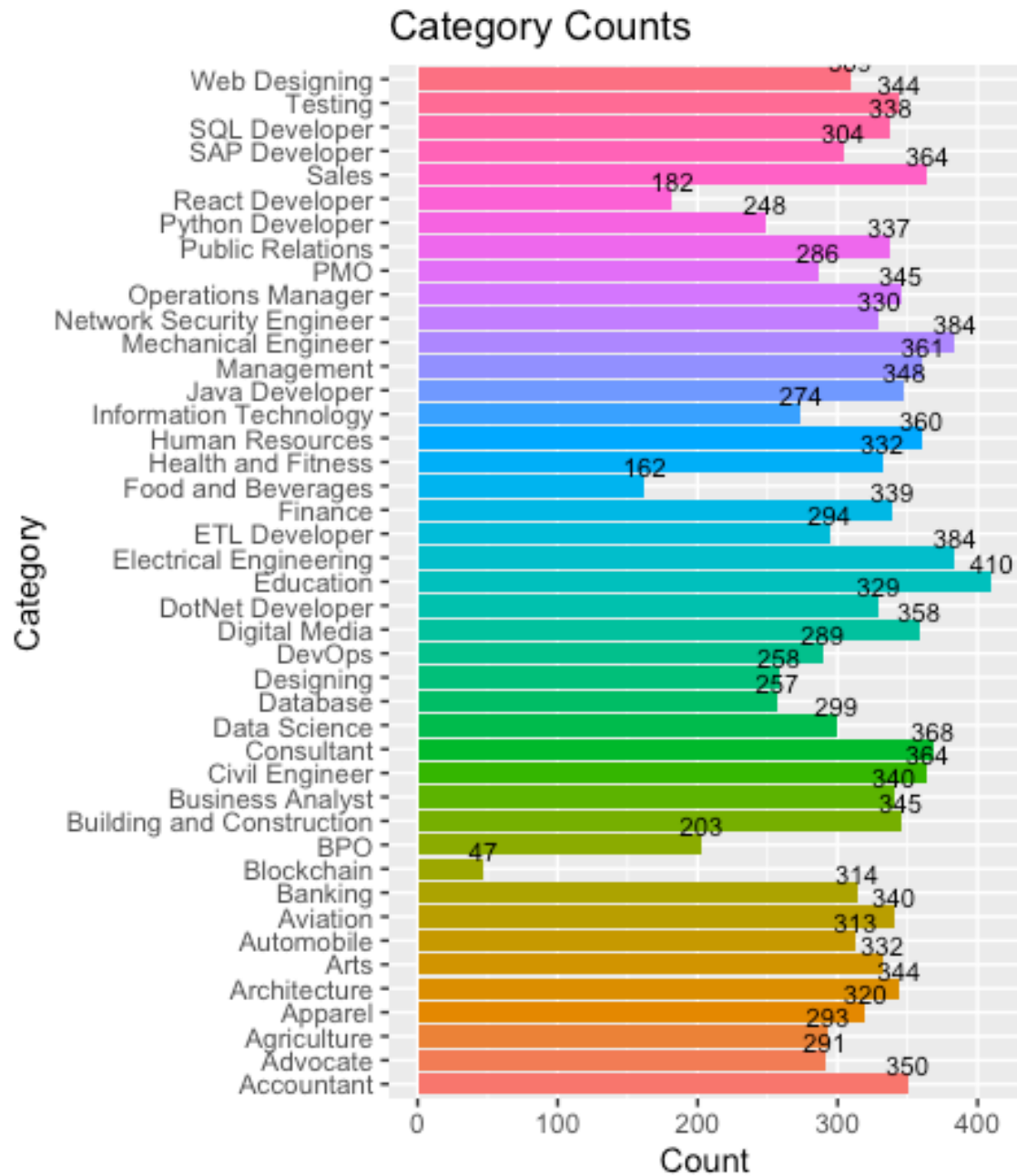


Figure 1. A bar chart outlining the number of resumes in the dataset for each job category.

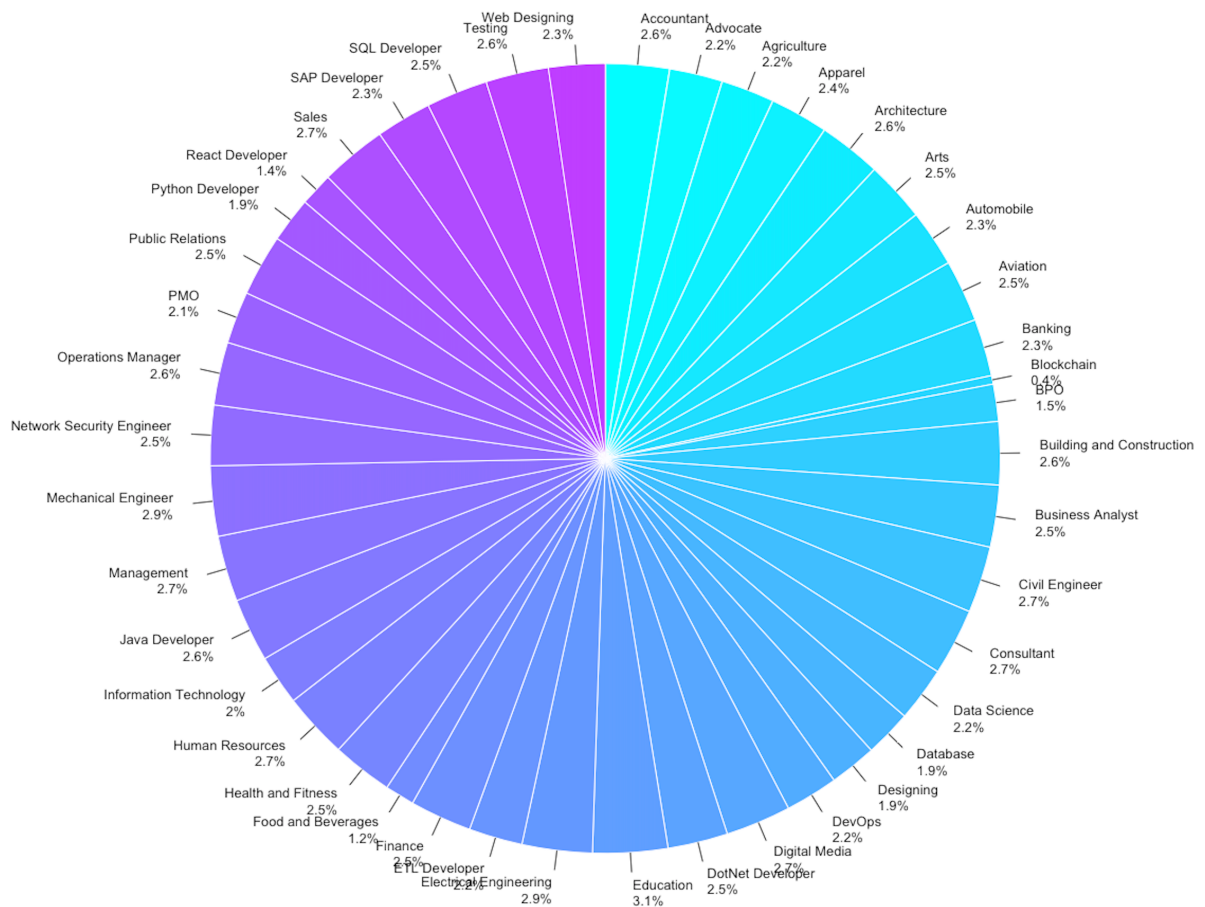


Figure 2. A Pie chart outlining the percentage of each job category in the resume data set.

A bar chart was created to display the absolute count of the resumes for each job category, which highlights that categories, such as Education and Mechanical Engineer are overrepresented in the dataset. Conversely, categories, such as Blockchain or Food and Beverages are severely underrepresented in the dataset (Fig. 1). A pie chart was also generated to represent the relative proportions of each category (Fig. 2). These visualizations offer clear views of how the dataset is distributed across various job titles and emphasize over and underrepresented categories.

The data instrumented in this study is freely accessible via Hugging Face and its comprehensive nature makes it an ideal resource for exploring the challenges and opportunities of applying AI in recruitment and resume categorization.

METHODS

Data Preprocessing

To begin the analysis of the resume dataset, preprocessing of textual data was completed. This step is essential in any Natural Language Processing (NLP) task because raw text data typically contains noise, or irrelevant and redundant elements. Noise can interfere with the analysis and model performance, so it must be handled before processing begins. The main goal of this preprocessing was to clean and standardize the text, so the entries could be more easily analyzed and compared.

The first step in this process was to convert all textual data to lowercase. This is a common technique in text normalization to ensure that words with and without capitalized letters will be treated identically. Without completing this step, the model could misinterpret, for instance, “Hello” and “hello” as two different words. This could lead to biased results, especially when examining word frequency.

The punctuation in the textual data was removed using Python’s *Natural Language Toolkit (NLTK)*. As punctuation does not typically add much information to text-based tasks, removing them allows the model to focus solely on the words.

Finally, stopwords, such as “the” or “and” were removed as they generally do not provide much extra information and may dilute the importance of more meaningful words. By preprocessing the textual data, more informative analyses can be completed.

Name Extraction

Next, to examine the textual data more effectively, Named Entity Recognition (NER) was used. NER is an NLP technique designed to locate and classify named entities in text, including people, organizations, or places. In this study, the names of the candidates were the focus. Spacy has a pre-trained NER model, which has been trained by a large amount of data and can recognize various types of named entities.

Entities labeled as “PERSON” were the most essential, which spaCy uses to detect mentions of an individual’s name. Names were successfully detected, and a column was added to the original dataset for further analysis. Once completed, to determine the accuracy of spaCy, the original dataset was compared to a census dataset.

Gender Prediction

Once the names were successfully extracted from the dataset, the next step was to predict the gender of the individuals identified in the text. Although gender cannot always be reliably inferred from names alone, many cultures have a strong associations between the first names of individuals and gender. This allows one to make educated predictions of gender based on first name.

The first name for each resume entry was selected based on a dataset containing a list of thousands of first names. The gender was then predicted using the same method, whether the first name was typically that of a male or female individual. Predicted gender was then added to the dataset as a separate column.

Ethnicity Prediction

Similar to gender prediction, the next step was to predict the ethnicity of the individuals based on their first names. This task was fairly challenging due to the fact that, like gender, ethnicity cannot always be reliably inferred from first names. However, some names have strong cultural or ethnic associations that can be used to make predictions. For this task, the Ethnicolr module was utilized to make predictions based on first name. The ethnicities that were predicted from the resume dataset are as follows:

- Germanic
- British
- Italian
- Jewish
- French
- Nordic
- Hispanic
- East European (generalized)

K-Nearest Neighbors Classification

For job category prediction, a k-Nearest Neighbors (k-NN) classifier was employed. This algorithm was chosen because it is simple, interpretable, and often effective in text classification tasks when the dataset is not too large. The model was trained using 80% of the data and tested on the remaining 20%. Evaluation of the model performance was carried out using standard metrics like accuracy, precision, recall, and F1-score.

RESULTS



Figure 3. Word Clouds detailing the most used words between the male and female resume selections.



Figure 4. Word Clouds detailing the most used words among certain ethnic populations.

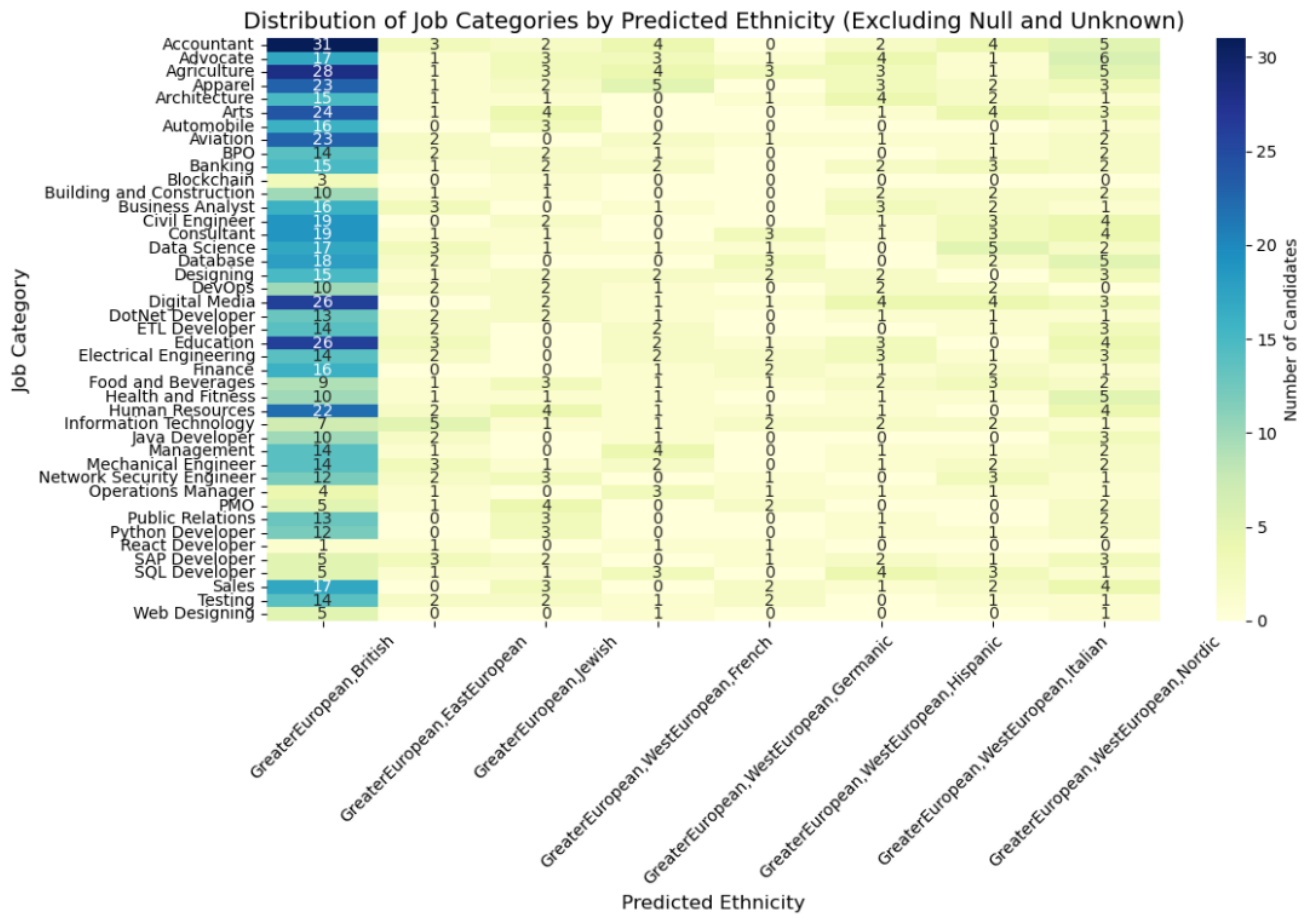


Figure 5. Heatmap detailing the distribution of job categories by predicted ethnicity.

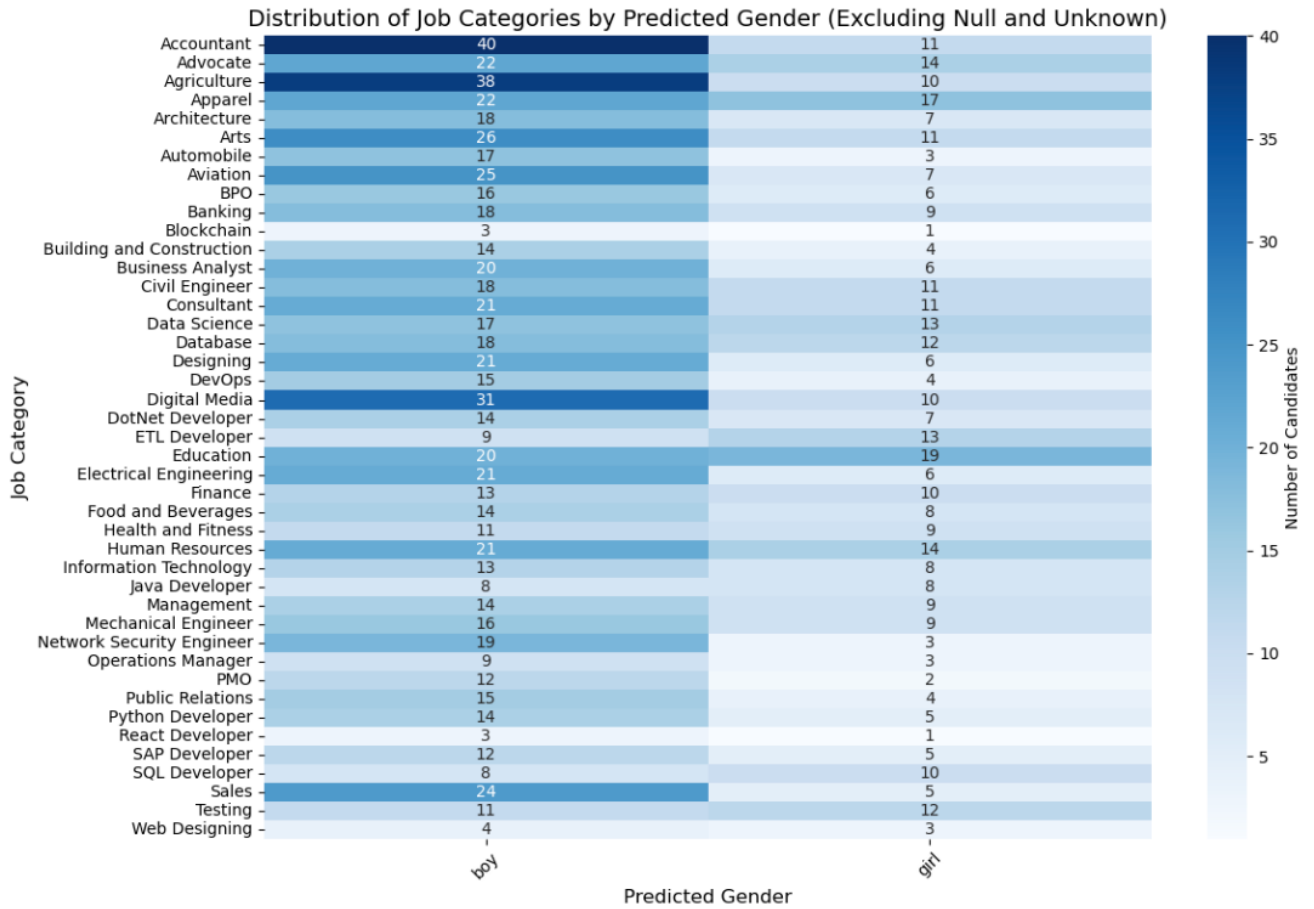


Figure 6. Heatmap detailing the distribution of job categories by predicted gender.

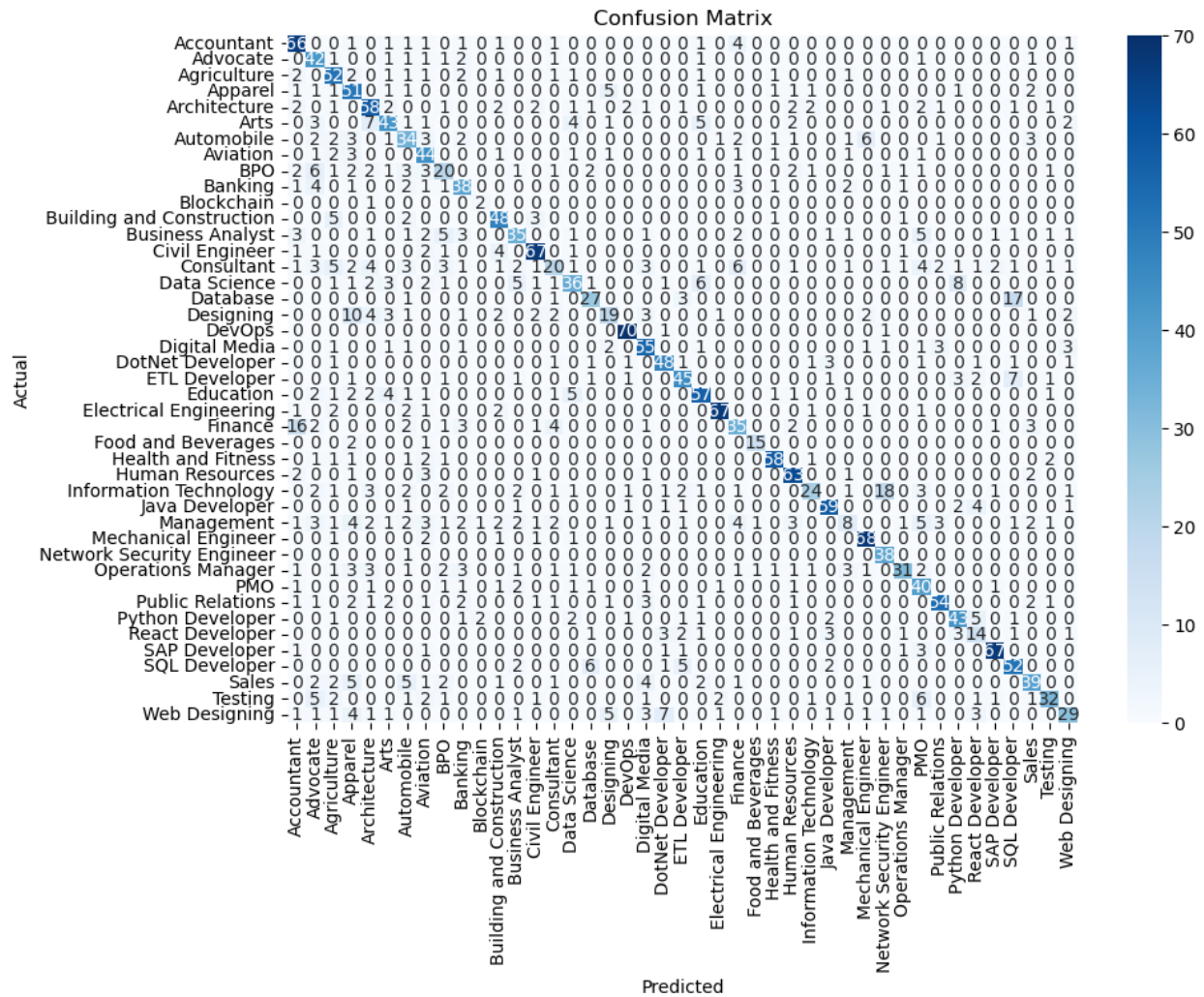


Figure 7. Confusion Matrix of k-Nearest Neighbor analysis outlining the predicted and actual values for each job category.

DISCUSSION

K-Nearest Neighbors Analysis

Given the results from the KNN model, it is evident that the model can successfully predict job category. The KNN analysis in this study predicts the category of a resume purely based on the textual content. The absence of historical hiring data in the prediction process means that the model is uninfluenced by past hiring decisions. This would be simple to replicate with a generic job description, outlining the necessary skills and qualifications for a specific job category. This could then be used to unbiasedly recommend candidates for specific jobs.

There is an advantage when not relying on historical hiring data. Specifically, it can prevent bias that might otherwise be introduced if the AI systems are trained on past hiring decisions. There is also a disadvantage, as the system in this study is heavily based on the textual features within the resumes. This means that the language patterns, phrasing, and types of qualifications that are emphasized on resumes are considered. This could choose one resume over another based on language, exhibiting implicit bias.

Gender and Ethnic Language

The top words for male and female candidates show clear differences, particularly in the emphasis on specific fields. Male candidates frequently highlight terms like “management” and “business,” while female candidates emphasize “data” and “design” more often. Similarly, different ethnic groups choose different words within their resumes. Western European and Nordic candidates emphasize “management” while British and Jewish candidates emphasize “business and “new” more frequently.

Avoiding Discrimination

By focusing on resume content and job qualifications—rather than historical hiring data or demographic characteristics—an AI-driven system could assist in improving fairness among candidates. For example, a model’s reliance on experience and skills as primary features could ensure that individuals who meet the job criteria are considered for roles that align with their expertise. This would be regardless of gender or ethnic background.

By removing historical data, reinforcement of the historical trends is also negated. Traditional hiring systems oftentimes perpetuate biases because past hiring decisions were influenced by human subjectivity, which could have been affected by bias.

To prevent this, the model should be trained with a diverse and balanced dataset that accurately reflects the full range of qualifications, skills, and experiences found across all gender and ethnic groups. Ongoing evaluation and testing should also be completely regularly to adjust for any potential disparities in the languages used across different demographic groups.

CONCLUSIONS

AI-driven resume screening systems have the potential to reduce human biases that may exist in traditional hiring practices. This is achieved by focusing on the qualifications and experiences of candidates rather than their demographic characteristics. In theory, AI can enable hiring decisions that are based purely on the content of resumes. In turn, ensuring that all applicants are evaluated on their abilities and qualifications rather than unconscious human biases.

However, it is crucial to recognize that these systems inherently come with risks. The data used to train AI models can inadvertently encode existing biases, specifically when it comes to linguistic and cultural differences between applicants. For example, male and female candidates may use different language or tones in their resumes. Similarly, ethnic groups may vary in how they describe their experiences. If historical hiring data is used to train these models, it can reinforce past biases, creating systems that favor certain groups over others.

To mitigate these risks, AI-resume screening systems should prioritize content-based features that assess applicants based on their qualifications, experiences, and skills. By doing so, fairer systems can be designed to assist in leveling the playing field. This would enable job candidates to be evaluated based on their true potential rather than unconscious or conscious biases.

Ultimately, a well-designed AI hiring system could create more equitable and transparent hiring processes that reduce the likelihood of discrimination.

DIRECTIONS FOR FUTURE WORK

There are several important directions for future work to improve the fairness and performance of machine learning models for resume categorization:

- **Data Diversification:** One of the most pressing needs is to diversify the dataset. Increasing the representation of underrepresented groups, such as females, non-European ethnicities, and less common job categories, will help reduce bias and improve model performance.
- **Bias Mitigation Techniques:** Future research should explore bias mitigation techniques, such as reweighting, oversampling underrepresented groups, or using fairness-aware algorithms. These methods can help reduce the disparate impact observed in gender and ethnicity predictions.
- **Model Enhancement:** Experimenting with other machine learning models, such as Random Forests or Support Vector Machines, could help improve prediction accuracy, particularly for categories with fewer training examples. Additionally, incorporating deep learning models, such as recurrent neural networks, might allow the model to better capture contextual information from the resume text.

- **Ethical Implications:** It is critical to evaluate the ethical implications of using machine learning for resume screening. Future work should consider how these models might be misused in recruitment.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Northwestern Data Science Master's Program faculty for providing me the opportunity to pursue this academic experience. The knowledge, skills, and guidance I have gained from this group of professors and advisors has helped me grow as a data scientist. Next, to my family, thank you for the unconditional encouragement, love, and belief in me. It has been a constant source of strength for me and I could not have done this without you all. A special shoutout to all of my closest friends who have been there for me in the challenging and rewarding moments throughout this program. You have kept me motivated and grounding throughout the entire process. To all my coworkers who have pushed me to be the best version of myself that I can, I thank you. I will miss you but am appreciative of your unwavering support for me. Lastly, anyone and everyone, whether directly or indirectly, that has supported me throughout this journey, I owe my gratitude. Thank you all.

DATA AVAILABILITY

The datasets used in this project are publicly available on Hugging Face or GitHub.

The Resume Dataset is available under the repository name InferencePrince555 under the following link:

<https://huggingface.co/datasets/InferencePrince555/Resume-Dataset>

The First Names dataset is available in the following GitHub repository:

<https://github.com/hadley/data-baby-names/blob/master/baby-names.csv>

CODE AVAILABILITY

The code written for this research is available for other academic and research purposes and will be made accessible from the author's GitHub Repository:

<https://github.com/jschapym/Chapman-Final-Project-12.2024>

This includes all code and documentation required to replicate these findings described in the study. For any questions regarding the code, please refer to the contact information found in the title page.

REFERENCES

- Angwin, J., Larson, J., Mattu, S., and Kirchner, L. 2016. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks." *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Binns, R. 2018. "Fairness in Machine Learning: Lessons from Political Philosophy." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. Association for Computing Machinery. <https://doi.org/10.1145/3173574.3174153>.
- California Legislative Information. 2020. "California Privacy Rights Act of 2020 (CPRA)." Retrieved from <https://leginfo.ca.gov>.
- Dastin, J. 2018. "Amazon Scrapped Secret AI Recruiting Tool That Showed Bias Against Women." *Reuters*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
- Gillespie, T. 2018. "Algorithmic Bias: The Politics of Artificial Intelligence in Recruitment and Hiring." In *The Ethics of AI and Robotics*, edited by Vincent C. Müller, 109–126. Springer. https://doi.org/10.1007/978-3-319-93951-4_8.
- Green, M. 2019. "Unconscious Bias in Recruitment: Why AI Can't Fix It Alone." *Harvard Business Review*, June 13. <https://hbr.org/2019/06/unconscious-bias-in-recruitment-why-ai-cant-fix-it-alone>.
- Hao, K. 2020. "How AI Is Changing the Recruitment Industry." *MIT Technology Review*, February 25. <https://www.technologyreview.com/2020/02/25/126329/how-ai-is-changing-the-recruitment-industry/>.
- Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2019. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 52 (1): 1–35. <https://doi.org/10.1145/3285029>.
- New York City Department of Consumer and Worker Protection. 2023. "Automated Employment Decision Tool Law." Retrieved from <https://www.nyc.gov>.
- O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
- Zafar, M. B., E. Valdano, and K. P. Gummadi. 2017. "Fairness in Machine Learning: Theoretical Aspects and Practical Concerns." *Journal of Machine Learning Research* 18 (1): 1–50. <http://www.jmlr.org/papers/volume18/17-367/17-367.pdf>.