CMPU-250 Research Project:

Gender Bias in Online Job Screenings

Maya Thompson, Zoe LaCheen-Fine, and Anh Duong

Fall 2025

1. Introduction and Data

Hiring algorithms are being increasingly used in candidate screening and selection processes. The utilization of hiring algorithms has been proposed to result in reduced time-to-hire, better ability for a company to assess soft skills, and more efficient filtering of candidates and talent pools, with the most significant areas of application being in the automation of screening and ranking in the recruitment and early-stage hiring processes (Zhang & Yencha, 2022). However, public opinions on algorithms used in job screening and hiring processes are notably divided. Although algorithms are often presented as fair and impartial, many caution that hiring algorithms may replicate discriminatory practices, inherit the bias of human decision-makers, or mirror biases in society (Zhang & Yencha, 2022).

Considering this background, we plan to examine gender bias in online job screenings, plus bias in specific industries of job listings. We want to investigate the possible presence of gender bias in the online job screening process across different industries. We also want to explore how gender-coded language found in job listings and resumes may influence hiring decisions. As women in STEM fields, we are acutely aware of the barriers that gender bias, whether subtle or implicit, may impose in hiring, advancement, and representation. The STEM field is a rapidly growing employment sector and one where women remain underrepresented. Furthermore, as college students, we are beginning to experience the benefits and harms of online job screenings. As individuals who hope to have jobs one day, we are interested in understanding the systems in place today that can both enhance and hinder our work experiences. When firms use algorithmic systems for screening, such as automated resume parsers, coding tests, or AI-assisted tools, there is a risk that these systems amplify or encode social biases. Our project, therefore, examines not only whether gender bias exists in online job screening but also how contextual features such as gender-coded words in job descriptions and industry differences influence candidate outcomes. By combining resumes (Kaggle, HireItPeople), screening systems (HackerRank, CodeSignal, ChatGPT), and analysis of job description text, we aim to provide evidence that speaks directly to questions of fairness in hiring algorithms.

1.1 Research Questions

We propose the following research questions:

1. Do male and female candidates with similar qualifications have different job screening outcomes?

2. Does gender bias in job screening vary across industries (specifically gender-dominated fields)?

3. Are gender-coded words (e.g., “collaborative” vs. “assertive”) in job descriptions associated with differences in the gender distribution of job applicants?

1.2 Dataset Construction

The dataset is constructed through web scraping of resumes available on Hire IT People. HireITPeople hosts thousands of applicant resumes for a wide range of fields in technology. For example, Java Developers/Architects Resumes, Web Developer Resumes, and Project Manager Resumes can be found. We focused on Java Developers/Architects for our research. The website originally had over 85,000 resumes but we decided to scrap 600 for time and efficiency. Each resume entry contains text under sections such as Summary, Technical Skills, and Professional Experience. After scraping, we preprocess the HTML to plain text and parse these sections to extract key variables. Specifically, we identify the applicant's years of experience, skills, number of past jobs, and industries worked in. As with the first dataset, we add a randomized "Gender" column to test our research question about bias.

1.3 Expected Findings

We hypothesize that there will not be a large bias based on gender alone, but algorithms' heavy reliance on other variables, such as gender-coded keywords in specific industries, may reflect implicit biases. We plan to use a regression model to evaluate whether hiring rates are different based on gender. A separate regression can evaluate industry-specific hiring decisions to see if industry-specific hiring preferences reflect gender bias.

1.4 Analysis of Data

a. Exploratory Data Analysis (EDA)

We began by summarizing the dataset using descriptive statistics and visualizations. This included examining distributions of years of experience, number of past jobs, and the frequency of technical skills across resumes. This step helped verify that the extracted features were consistent and allowed us to identify anomalies, missing information, or formatting issues before applying the scoring function.

b. Resume Scoring and Outcome Generation

Rather than using AI-based screening tools, we developed a transparent scoring function to simulate automated hiring assessments. Each resume was evaluated based on a weighted combination of three factors: semantic similarity between the resume and a job description, years of experience, and number of past jobs. These scores served as our outcome variable, representing how an automated system might rank applicants in a real-world screening process. The scoring system allowed us to systematically examine the relationship between resume content and assigned gender labels without relying on black-box AI models.

c. Feature Analysis and Gender-Coding

We analyzed how specific resume characteristics, particularly the presence of gender-coded language, influenced scoring outcomes. Using methods inspired by Gaucher, Friesen, and Kay (2011), we identified masculine- and feminine-coded words in resumes and examined whether resumes containing these terms systematically received higher or lower scores. This allowed us to test whether linguistic patterns associated with gender could create disparities in ranking outcomes, even when gender was not explicitly used in the scoring function.

1.5 Outcome and Key Variables

Our primary outcome variable is the resume score, which reflects whether a candidate would be ranked higher in a simulated automated screening process. Key explanatory variables include technical skills, years of experience, number of past jobs, assigned gender, and counts of masculine and feminine-coded words. These variables form the basis of our analysis on how gender and language interact to influence resume rankings.

1.6 Data Cleaning and Preparation

To prepare the data, we scraped 600 of the 85,000 resumes from HireItPeople.com under the Java Developers/Architects Resumes page. We created the resume.csv file by opening each link that holds a resume and then added the text to the CSV. Once we had a CSV containing the URL link and the long resume text, we were able to begin separating the different variables into separate columns.

To clean the data, we separated the resume data into more specific column categories. First, we separated the data by “summary,” “skill level,” and “technical skills.” Then, we started extracting certain details such as number of past jobs and number of years of experience.

We removed some filler information, such as repetitive language, and excluded resumes with the “professional summary” section repeated (3 entries). We randomly assigned gender (male/female) to each entry. We also added columns for years of experience and number of previous jobs held, which were pulled from the resume descriptions.

1.7 Gender-Coded Words

We used Gender Decoder, which was inspired by a research paper written by Danielle Gaucher, Justin Friesen, and Aaron C. Kay (2011). In this paper, *Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality*, researchers showed job adverts that included different kinds of gender-coded language to men and women and recorded how appealing the jobs seemed and how much the participants felt that they 'belonged' in that occupation.

2. Methods

Our resume scoring algorithm combined three quantitative indicators of job fit: (1) semantic similarity between each resume and the job description, (2) the applicant’s years of experience, and (3) the number of previous jobs listed. The three inputs were normalized and combined using a weighted linear formula, with semantic similarity contributing the most (70%), followed by years of experience (20%) and job count (10%). These weights were chosen to prioritize content relevance while still incorporating basic measures of experience.

Semantic similarity was calculated using a TF-IDF Vectorizer, which measured if applicants' resumes matched against a set of keywords that suited the resume description. TF-IDF similarity was chosen because it is simple, transparent, and interpretable. It provides a baseline measure of how well the content of each resume aligns with the language and requirements of the job description.

Experience features were normalized using min-max scaling so that each feature contributed proportionally to the final score regardless of its original range. Years of experience and job count were selected as lightweight proxies for career depth and stability.

In the analysis, for each applicant, the algorithm computed:

- Overall score = (0.70 * semantic similarity) + (0.20 * years of experience) + (0.10 * number of past jobs)

The final rankings were assigned by sorting applicants in descending order of their overall score. Their individual scores, based on semantic similarity alone, and demographic features were preserved for descriptive statistics, but were not considered in the final ranking/scoring.

2.1 Revised Algorithm

Our revised resume scoring algorithm combined three quantitative indicators of job fit: (1) hard skill keywords, split into required and preferred skills, (2) soft-kill keywords, and (3) job experience, measured using years of experience and number of prior jobs. Rather than using the TF-IDF vectorizer, resumes were scored by counting whether each keyword appeared at least once in the resume text, then normalizing each component using min-max scaling and combining them with a weighted linear formula. Resume text was cleaned by lowercasing, normalizing punctuation, and standardizing programming-language tokens (e.g., C#). Keyword-matching used a case-insensitive pattern with flexible whitespace and hyphen handling to detect whether a keyword appeared in resume text. For each resume, we computed three keyword-based counts: (1) the number of required hard-skill keywords matched, (2) the number of preferred hard-skill keywords matched, and (3) the number of soft-skill keywords matched.

Years of experience and number of past jobs were extracted from numeric columns and normalized using min-max scaling so that each feature contributed proportionally to the final score regardless of its original range. Years of experience and job count were selected as lightweight proxies for career depth and stability. These were combined into a continuous experience score using a weighted average (60% years, 40% job count). In addition, a binary preferred experience bonus was computed to match the preferred years of experience in the job description. Applicants received a value of 1 if their years of experience were greater than or equal to the preferred number of years of experience, otherwise 0. This bonus contributed to the preferred hard-skill component.

After normalization, the algorithm computed:

- Experience score
  - Score = (0.60 * years of experience) + (0.40 * number of jobs)
- Initial hard skills + preferred experience
  - Score = (0.75 * number of keywords matched) +
    - if years of experience >= preferred, then added bonus of (0.25 * years of experience)
- Normalized hard skill score (required vs preferred)
  - score = (0.50 * required hard skills) + (0.25 * preferred hard skills)
- Overall score:
  - Final score = (0.50 * hard) + (0.20 * soft) + (0.30 * experience)

2.2 Revised Job Descriptions

After analyzing our first method and results, we created three new job descriptions for the revised algorithm. We took an existing Java developer job description from a hiring website and then created two new job descriptions with injected masculine and feminine language. The original job description was taken from CFS Staffing in which we used the responsibilities and preferred skills that were provided. Next, we took some of the gender-coded language and made two new descriptions: one that was heavily masculine and the other heavily feminine. Our goal was to see if the job description would have any influence on the outcome of the algorithm.

3. Results

3.1 Descriptive Profile of the Applicant Pool

Across the entire dataset, applicants show a relatively consistent pattern of career experience and resume structure.

3.1.1 Years of Experience

Years of experience form an approximately normal distribution centered around seven years. Most applicants fall between four and ten years of experience, which is typical for mid-career hiring pools. The distribution appears slightly different across genders when those genders are assigned randomly: women show a modest left skew, while men show a slight right skew. This difference, however, is small and does not indicate any meaningful variation in qualifications at this stage of analysis.

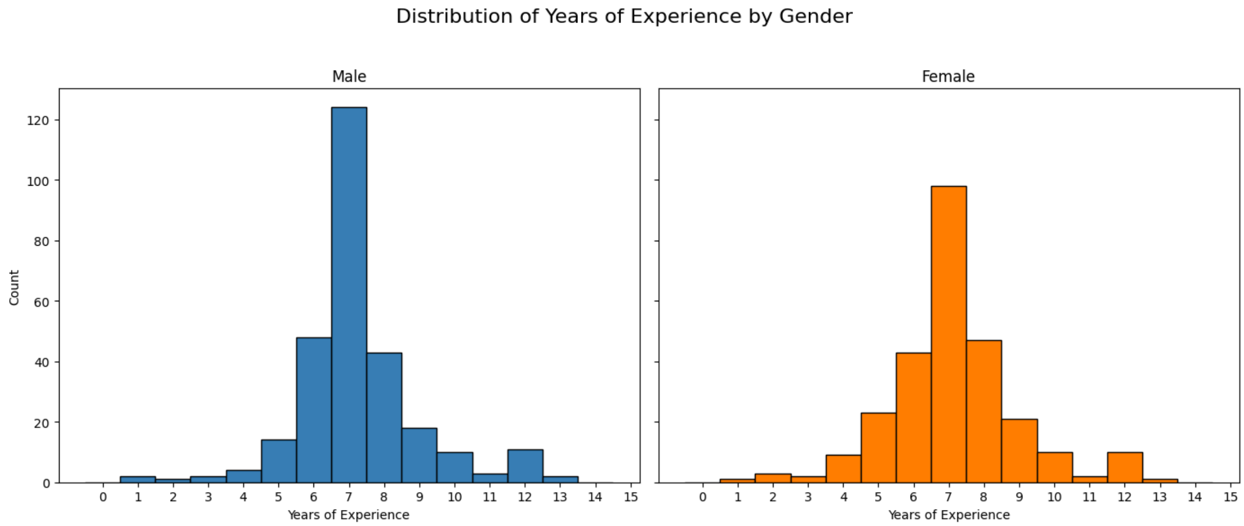

Figure 1*. Distribution of Years of Experience by Gender*

3.1.2 Number of Past Jobs

The number of past jobs follows a right-skewed distribution, with most candidates having held approximately five positions. Very few candidates list more than ten prior roles. This pattern is consistent across genders: regardless of assignment method, men and women display nearly identical distributions in job mobility. This suggests that the dataset does not contain systematic gender differences in employment stability or career transitions.

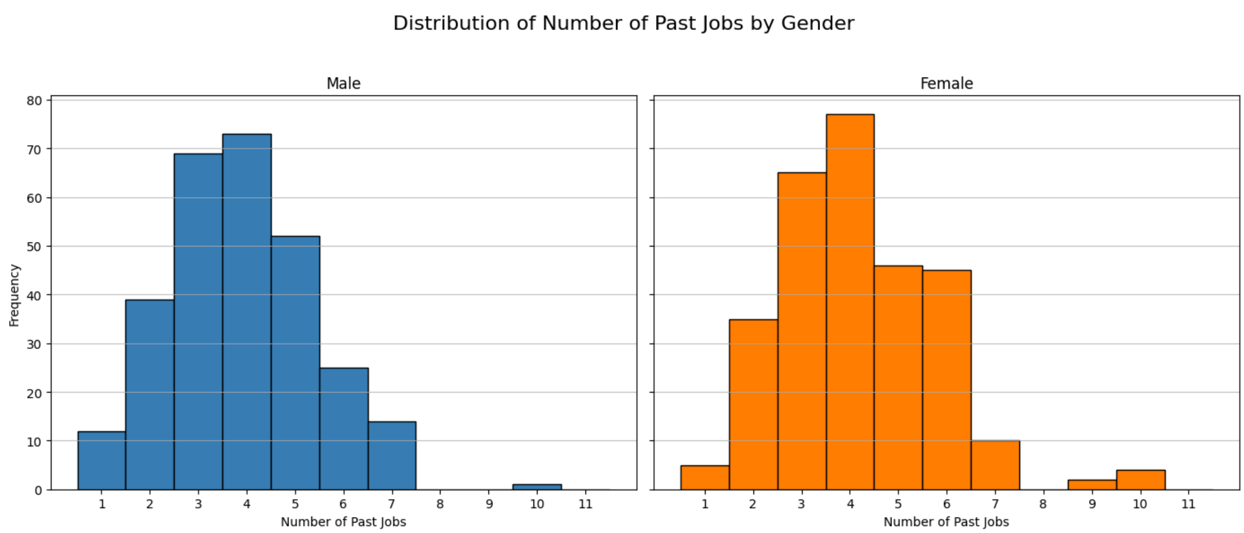

Figure 2*. Distribution of Number of Past Jobs by Gender*

3.1.3 Number of Masculine coded and Feminine coded words

Finally, we computed the number of masculine- and feminine-coded words in each resume. Masculine-coded words were more common across the dataset as a whole, and this pattern remained stable regardless of gender assignment method. Even among applicants randomly labeled as female, masculine-coded phrasing appeared frequently, indicating that linguistic style is not evenly distributed across the applicant pool.

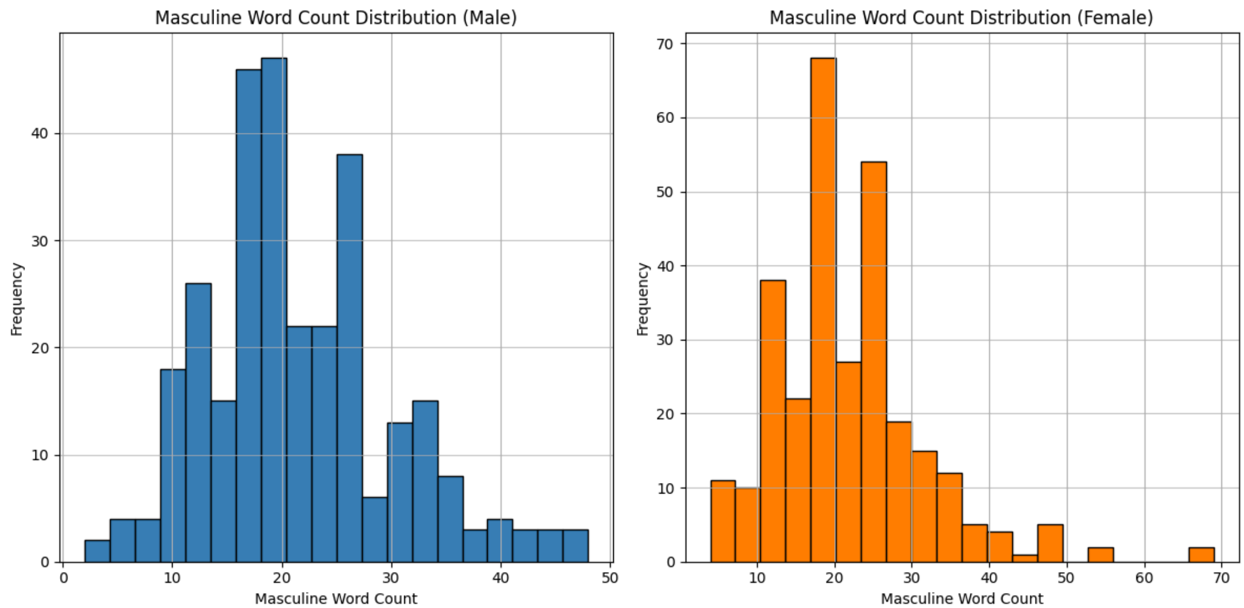

***Figure 3**. Distribution of Number of Masculine coded words by Gender*

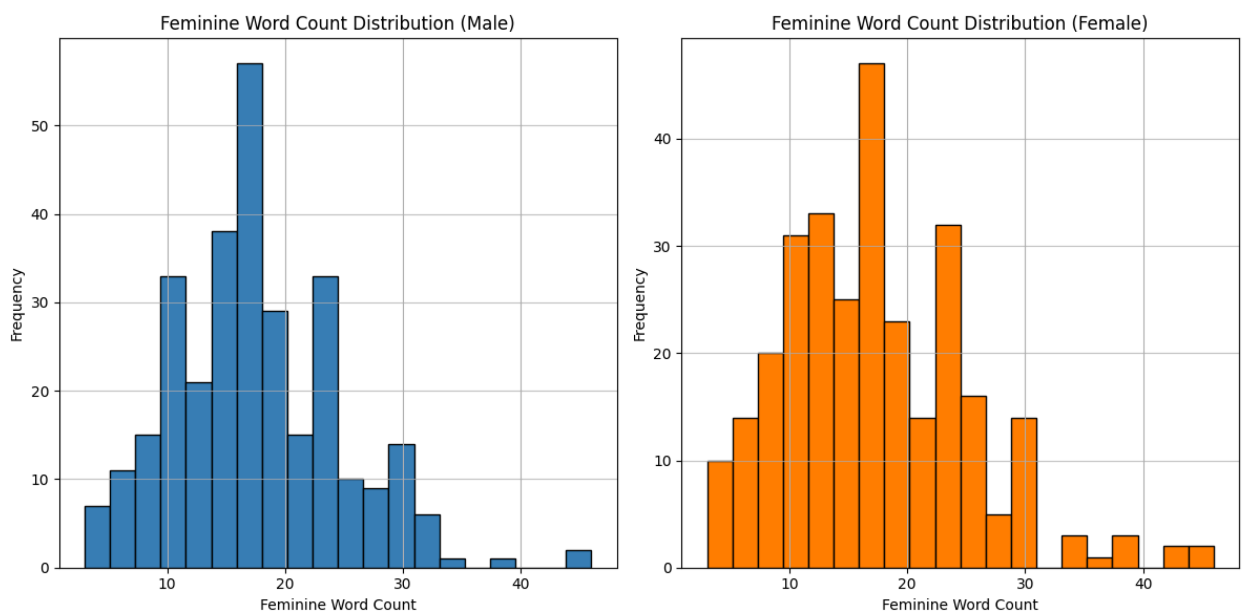

***Figure 4**. Distribution of Number of Feminine coded words by Gender*

3.2 Algorithmic Rankings Under Random Gender Assignment

To evaluate whether the resume-scoring algorithm produced different outcomes for men and women when gender was not correlated with resume content, we first randomly assigned gender to all applicants without an existing label, fixing the percentage of male and female to be roughly 50% to have a balanced dataset. When the algorithm ranked applicants by the weighted score combining semantic similarity, years of experience, and number of past jobs, the top 50% of applicants contained proportions of men and women that were similar to the overall distribution. In other words, when gender was independent of resume content, the algorithm did not disproportionately favor one group.

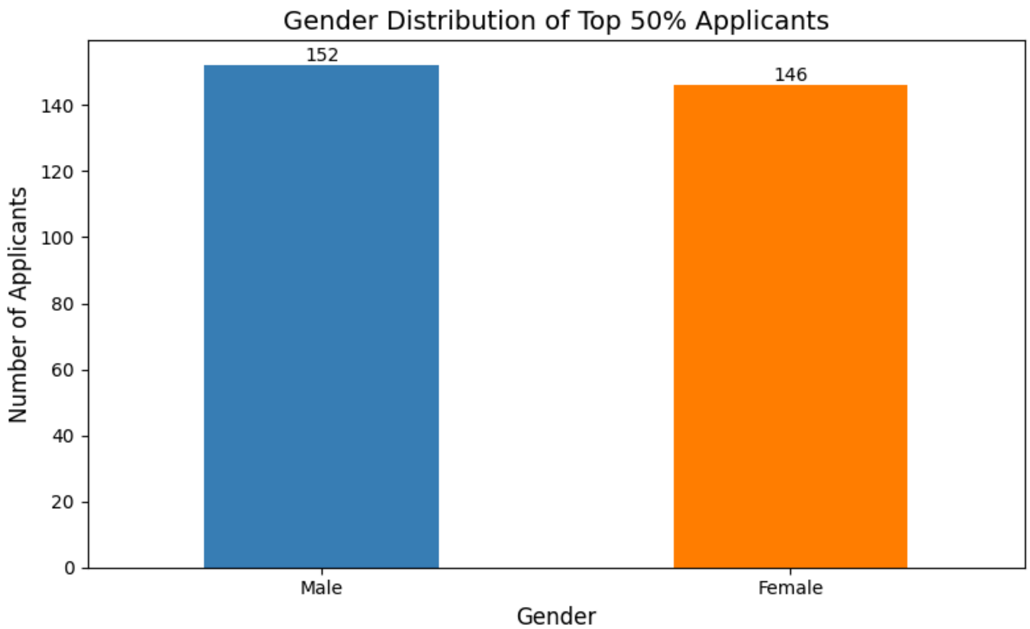

***Figure 5**. Gender Distribution of Top 50% Applicants*

This result provides an important baseline: if gender has no structural relationship to the features the algorithm rewards, then the ranking outcomes appear unbiased.

3.3 Algorithmic Rankings Under Text-Based Gender Assignment

The interpretation changes substantially when gender is inferred from linguistic signatures, specifically, the relative frequency of masculine- and feminine-coded words. Under this condition, the dataset becomes uneven (*Figure 6)*. A noticeably larger share of resumes are classified as male due to their higher counts of masculine language. Feminine-coded resumes appear less frequently, suggesting that applicants in this dataset, regardless of actual gender identity, tend to describe their skills and accomplishments using more stereotypically masculine phrasing. This imbalance has downstream consequences. Because masculine-coded resumes tend to be longer, contain more technical verbs, and score slightly higher in TF-IDF similarity to the job description, their semantic similarity scores are, on average, higher. Moreover, these resumes often list more detailed tasks and responsibilities, inflating their number of past jobs and boosting their overall score. Thus, although gender labels themselves never enter the scoring formula, features correlated with masculine-coded language do.

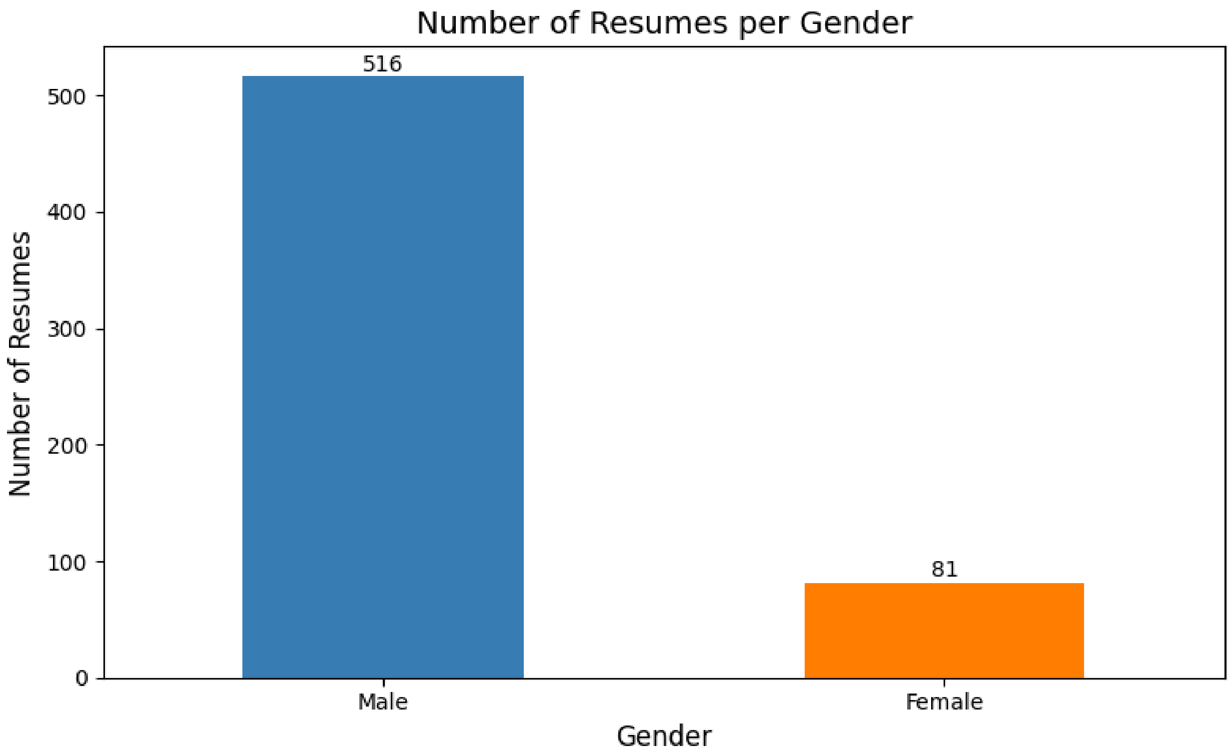

***Figure 6**. Distribution of Resumes by Gender using Linguistic Approach*

When we examine the top 50% of algorithmic rankings under this condition, the difference becomes pronounced (*Figure 7)*. Male-coded resumes occupy a disproportionately larger share of the top-ranking applicants, while female-coded resumes appear far less frequently. This shift is not due to algorithmic bias toward gender, but rather reflects the way linguistic style interacts with the scoring features, particularly semantic similarity, which was weighted most heavily.

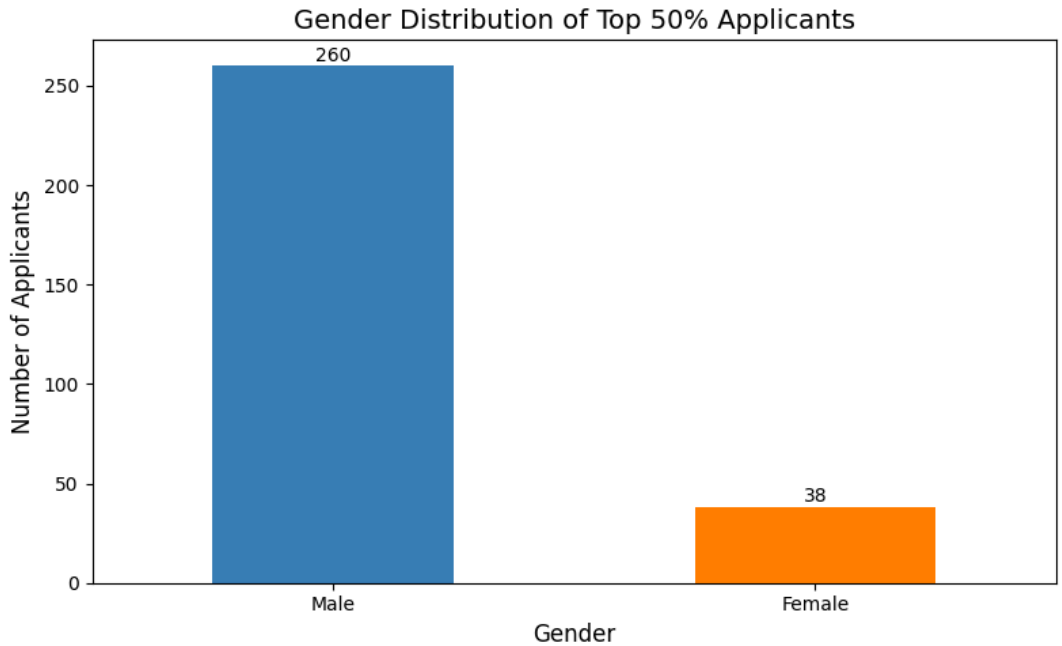

Figure 7*. Gender Distribution of Top 50% Applicants using Linguistic Approach*

3.4 Comparison Between the Two Gender Conditions

Juxtaposing the two gender-assignment methods highlights a crucial insight: **the perceived fairness of the resume-screening algorithm depends strongly on how gender is defined in the dataset**. Under random assignment, the rankings appear balanced and gender-neutral. Under a text-based assignment, they appear skewed, with male-coded resumes overrepresented among top performers. Both results are technically “correct” given the assumptions but they tell very different stories. This contrast suggests that disparities emerge not because the algorithm directly encodes gender bias, but because certain linguistic patterns, more frequently associated with masculine-coded writing, align more strongly with the algorithm’s scoring dimensions. The fact that masculine-coded resumes tend to be longer, more technical, and more detailed means they naturally accumulate higher similarity scores and, in turn, higher final rankings.

3.5 Comparison between Job Descriptions with new Weights

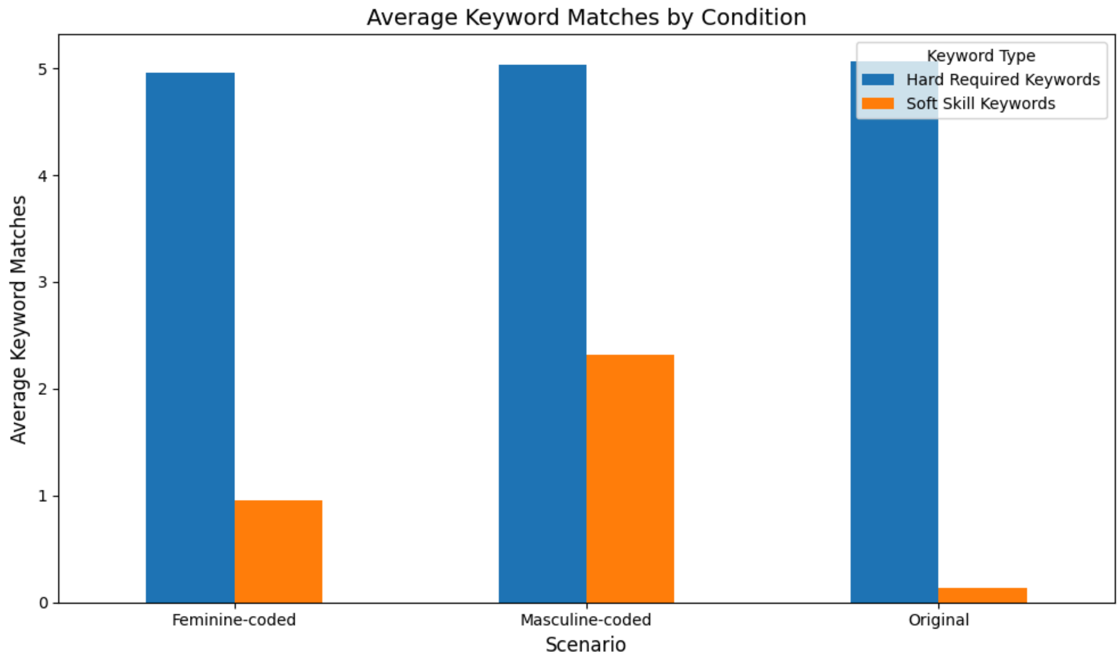

Figure 8. *Distribution of Keyword Matches by Condition*

The average keyword matches in the hard skill category did not significantly differ by job description conditions. However, we observed greater variation in the average soft skill keyword matches across conditions. The masculine-coded job description yielded the highest average soft skill matches, while the feminine-coded and original job descriptions had significantly fewer soft skill matches on average. These results suggest that while required hard skills may stay constant across job descriptions, the soft skill requirements may change and therefore may affect applicants' chances at being hired differently.

3.5.1 Original Job Description

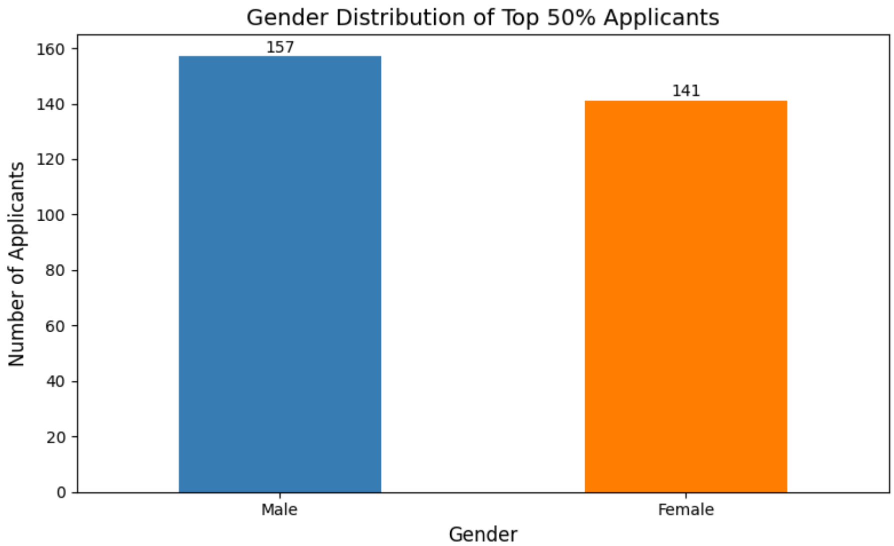

Figure 9. *Gender Distribution of Top 50% Applicants (Original Job Description)*

After running the new algorithm on the original, unchanged job description, there are 157 male applicant resumes and 141 female applicant resumes among the top 50% applicants. This shows that more males were better candidates than females for the java developer job description we found. There were more keyword matches and experience levels matched between the job description and the male resumes. This figure relates to our original hypothesis that there is a gender bias for job applications.

3.5.2 Injected Masculine

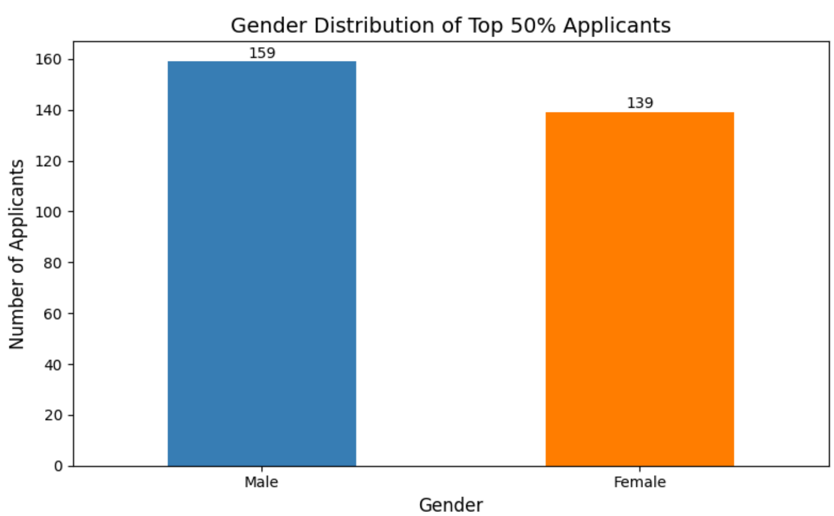

Figure 10. *Gender Distribution of Top 50% Applicants (Injected Masculine words)*

After running the new algorithm on the male gendered language injected job description. There were 159 male applicant resumes and 139 female applicant resumes. This shows that when there are more masculine-coded words present in a job description, more male resume applicants will match using the algorithm we created. This visualizes that there were more keyword matches and experience levels matched between the masculine job description and the male resumes.

3.5.3 Injected Feminine

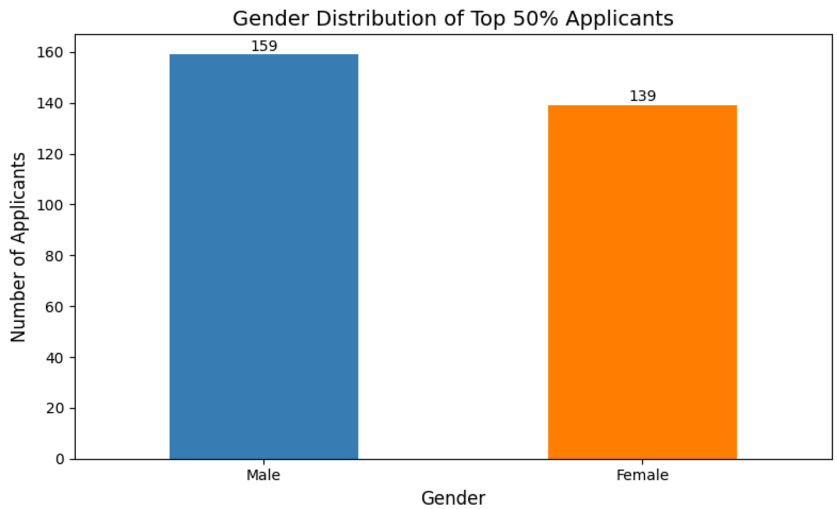

Figure 11. *Gender Distribution of Top 50% Applicants (Injected Feminine Words)*

After running the new algorithm on the female gendered language injected job description. There were 159 male applicant resumes and 139 female applicant resumes. This visualizes that even though there should be more keyword matches and experience levels between the job description and the female resume applicants, the male resume applicants still had a higher matching rate. This could be related to the types of words that are found in resumes.

4. Discussion

Our study examined whether gender bias appears in automated resume screening systems and how different ways of operationalizing gender affect observed outcomes. Across our analyses, we find that the appearance of bias depends less on gender labels alone and more on how linguistic features and job description wording interact with the scoring mechanism.

4.1 Gender-Coded Language and Emergent Bias

When gender was assigned randomly and held independent of resume content, the resume-scoring algorithm produced nearly equal proportions of male and female applicants in the top-ranked group. This suggests that the scoring function—based on semantic similarity, years of experience, and number of past jobs—does not generate gender disparities on its own when gender is not correlated with resume features. Under this condition, the model behaves as a neutral baseline. This result supports our hypothesis that gender alone does not drive large differences in outcomes.

When gender was inferred using gender-coded language, the composition of the applicant pool changed substantially. Masculine-coded resumes appeared more frequently than feminine-coded resumes, reflecting patterns commonly observed in technical fields where resumes emphasize assertiveness, independence, and technical expertise. These patterns are better understood as features of occupational norms and resume-writing conventions rather than indicators of individual gender identity.

4.2 Consequences for Algorithmic Outcomes

These differences in resume composition had direct consequences for ranking outcomes. Masculine-coded resumes tended to be longer and contained more technical detail. Because semantic similarity carried the greatest weight in the scoring function, these resumes systematically received higher scores. As a result, male-coded resumes were overrepresented among top-ranked applicants, producing an apparent gender disparity even though gender was not an explicit input to the algorithm. This demonstrates how correlated input features can produce unequal outcomes without overtly discriminatory rules.

4.3 Job Description Language and Skill Matching

Analysis of keyword matching revealed that hard skill matches did not differ meaningfully across job description conditions, suggesting that core technical requirements remained stable. In contrast, soft skill matches varied across conditions. The masculine-coded job description produced the highest average number of soft skill matches, while the feminine-coded and original descriptions produced fewer. This indicates that changes in job description language primarily affect soft skill alignment, which in turn influences overall scores and applicant rankings.

4.4 Relevance to Broader Literature

Our findings show that gender-coded language in job descriptions affects algorithmic rankings, even when hard skill requirements remain constant. While hard skill matches did not differ across conditions, soft skill matches were highest for the masculine-coded job description, giving male-coded resumes an advantage. This supports existing research showing that linguistic patterns in job descriptions influence hiring outcomes (Gaucher et al., 2011). Our results also highlight that even in a simplified, transparent scoring system, disparities emerge based on word choice rather than explicit gender indicators. This underscores the broader risk in automated hiring systems: proprietary or black-box algorithms are likely to amplify these language-driven biases. Practitioners should therefore consider not just which variables are included in an algorithm, but also how the wording of job descriptions and occupational norms shape the features that determine applicant rankings.

4.5 Conclusions

These findings show that fairness evaluations are sensitive to how gender is defined, represented, and measured in the study, and how linguistic features enter the model. When gender is independent of resume content, outcomes appear balanced. When gender is linked to language patterns that are unevenly distributed across resumes and job descriptions, disparities emerge. Importantly, these disparities arise from the interaction between language conventions and the scoring criteria, not from explicit gender-based decision rules. Overall, our results support the view that algorithmic bias in hiring systems often reflects structural and linguistic patterns embedded in the data. Removing explicit gender indicators is not sufficient to prevent unequal outcomes. Instead, attention must be paid to how resume-writing norms, job description wording, and feature weighting shape who is favored by automated screening systems.

4.6 Resume Scoring Function Limitations

The resume scoring/ranking function looked at a weighted combination of three factors: the semantic similarity between the applicant's resume text and the job description, the number of years of experience, and how many of the past jobs the applicant had. While this approach was easy to understand and implement, it had a few key limitations.

First, the initial semantic similarity between the resume text and job description was calculated using a TFIDF Vectorizer, which measures if keywords overlap. However, this scoring approach doesn't capture the semantic context or meaning of the words, so a resume that repeats the same words as the job description may score high even if the candidate's experience is weak. At the same time, someone who describes the right skills using different wording may score lower. We attempted to resolve some of these issues using a keyword-matching approach for our revised algorithm, which differentially weighted hard skills (with a distinction between required and preferred skills) and soft skills to determine overall 'fit' of a candidate to the job description. Nonetheless, we were unable to match our weights precisely to any existing hiring criteria, as we could not find any descriptions of how existing hiring algorithms calculate scores for applicants.

Second, the years-of-experience and job-count features are very crude. They assume that "more years" and "more jobs" always mean "more qualified," which is not always true. These features can also indirectly favor older candidates or penalize people with career gaps or fewer roles, even if they are strong fits for the job.

Third, the weights used for the scoring function are arbitrary (0.70, 0.20, and 0.10). They were not learned from data or validated in any way, and instead represented a programming decision, as we assumed the most influential aspect of an application would be its match to the job description, with years of experience and number of past jobs being less significant deciding factors, but still considered.

Finally, because everything is normalized across the current dataset, individual scores depend on who else is in the applicant pool. Adding or removing resumes would change the min-max scores and thereby change everyone's scores, making the results less consistent. Overall, this method works as a simple baseline for ranking resumes, but it should not be used as a hiring tool because it misses the true semantic understanding of keyword phrases, relies on oversimplified numeric features, and lacks validation. Improving it would require better semantic models, explicit skill extraction, and fairness checks.

4.7 Data Reliability and Validity

There are multiple things to consider pertaining to the reliability and validity of our data and analytical approach. Scraping resume information from Hire IT People was done as systematically as we thought possible. While the resumes all have the same or similar information, each resume varies in length, format, detail, and level of description. We scraped the entirety of each resume and then weeded out information. While this made sense and was feasible for us to do, it does result in some reliability issues when extracting the data. The validity of our data should also be heavily considered. Due to the fact that we randomly assigned gender labels, our data is not directly related to real-world scenarios or examples. While we were able to use this variable to help analyze results, we can also only take our results with so much truth. As a result, our analysis looks at how an automated system responds to a labeled gender signal, not necessarily how actual employers may infer gender from names, writing style, or career trajectories. Overall, our methods allow us to examine potential mechanisms of algorithmic bias, but our analysis must be interpreted cautiously given the structured nature of our dataset and the limitations of our data.

4.8 Limitations for Generalization

Furthermore, the data we used limits any inferential conclusions or predictions for new data we could make. First, since our dataset was scraped from a single public online repository (Hire IT People), it is not fully representative of the broader applicant population. We took a handful of resumes under Java Developers/Architects. This means we looked at only a small subset of resumes that are out there to analyze. In addition, many of these resumes are from high-achieving, highly technical professionals. This restricts our ability to make inferences about other groups of resumes under different job descriptions. We cannot make predictions for other fields, like non-STEM fields or entry-level positions. Second, the fact that we made gender random means that we created dummy data for us to use. This means that we would not really be able to make meaningful inferences about how the genders are treated. For example, gender may correlate with differences in experience patterns, skill domains, award types, or industry representation in actual labor markets, which our data does not take into account. Because of these limitations, our data and results should be interpreted as the potential presence of bias, rather than as predictive tools or conclusions that can be applied to new or real-world applicant data.

4.8.1 Revised Algorithm

In regards to our revised algorithm, it should be noted that we were unable to find a public algorithm from any hiring companies to base our algorithm on. The existing literature suggests that the fields we included (hard skills, soft skills, experience) are considered in job screening algorithms, but it is unclear how these components are weighted in reality. This means that our revised algorithm is arbitrary and likely does not reflect any real-world applications.

4.9 Ethical Considerations and Societal Implications

Our project raises ethical concerns and larger societal implications because it involves scraping resume data and simulating algorithmic hiring decisions. The fact that we used resumes from real applicants and personal information brings up privacy and data-use concerns. It is unknown whether or not the individuals who are connected to the resumes would want their personal information to be used for data research, such as our project. Ethical risks are also present when analyzing gender bias using randomly assigned gender labels, as there is a possibility of reinforcing stereotypes or oversimplifying the complexities of real gender identity. Our project does not accurately replicate what the real-world hiring process is like and how hiring algorithms or companies view or infer gender. One societal implication of our project is that if hiring algorithms do in fact have a gender bias, they could unfairly limit job opportunities for women. This means that biased algorithms could contribute to long-term inequalities in the job market. While our project has larger issues and implications, it also yields positive outcomes that help reduce bias in the hiring process, particularly for women, who continue to face challenges in this area.

5. Citations

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of Personality and Social Psychology, 101(1), 109–128. https://doi.org/10.1037/a0022530

Zhang, Lixuan, and Christopher Yencha. "Examining perceptions towards hiring algorithms." *Technology in Society* 68 (2022): 101848. https://www.sciencedirect.com/science/article/pii/S0160791X21003237#abs0015

6. Appendix

Masculine prefixes	'active', 'adventurous', 'aggress', 'ambitio', 'analy', 'assert', 'athlet', 'autonom', 'battle', 'boast', 'challeng', 'champion', 'compet', 'confident', 'courag', 'decid', 'decision', 'decisive', 'defend', 'determin', 'domina', 'dominant', 'driven', 'fearless', 'fight', 'force', 'greedy', 'head-strong', 'headstrong', 'hierarch', 'hostil', 'impulsive', 'independen', 'individual', 'intellect', 'lead', 'logic', 'objective', 'opinion', 'outspoken', 'persist', 'principle', 'reckless', 'self-confiden', 'self-relian', 'self-sufficien', 'selfconfiden', 'selfrelian', 'selfsufficien', 'stubborn', 'superior', 'unreasonab'
Feminine prefixes	'agree', 'affectionate', 'child', 'cheer', 'collab', 'commit', 'communal', 'compassion', 'connect', 'considerate', 'cooperat', 'co-operat', 'depend', 'emotiona', 'empath', 'feel', 'flatterable', 'gentle', 'honest', 'interpersonal', 'interdependen', 'interpersona', 'inter-personal', 'inter-dependen', 'inter-persona', 'kind', 'kinship', 'loyal', 'modesty', 'nag', 'nurtur', 'pleasant', 'polite', 'quiet', 'respon', 'sensitiv', 'submissive', 'support', 'sympath', 'tender', 'together', 'trust', 'understand', 'warm', 'whin', 'enthusias', 'inclusive', 'yield', 'share', 'sharin'