Measuring Algorithmic Bias in Job Recommender Systems: An Audit Study Approach

Shuo Zhang *

November 7, 2021
(Latest Version Here)

Abstract

This paper investigates gender bias in job recommender systems. By conducting an algorithm audit in four Chinese job boards, I find that gender-specific jobs, which are only displayed to one gender, account for 9.72% of the total recommended jobs to identical male and female applicants. Gender-specific jobs differ in both the job's explicit quality and the words used in job descriptions: Compared to jobs that are only recommended to men, only-to-women jobs propose lower wages, request fewer years of working experience, are more likely to require literacy skills and administrative skills, and tend to contain words related to feminine personality, which reflect gender stereotypes in the workplace. Item-based collaborative filtering, content-based recommendation algorithms and the hiring agents' behaviors incorporated in job recommender systems are the possible drivers of the gender bias in job recommendations.

Keywords: *Recommender System, Algorithm, Gender, Job Platform*

JEL Codes: C93, J71, J16, O33, M50

*Department of Economics, University of California, Santa Barbara, CA 93106. E-mail: shuo_zhang@ucsb.edu. I am deeply indebted to my advisor Peter Kuhn for his continued guidance, encouragement, and support. I am grateful to Kelly Bedard, Clément de Chaisemartin, Heather Royer, and Catherine Weinberger for their helpful comments and suggestions. This study was approved under UCSB IRB No.17-20-0451 and is in the AEA RCT Registry under AEARCTR-0006101. All errors are my own.

1 Introduction

With the rapid development of the Internet, the explosive growth of information makes it increasingly challenging for people to process a huge amount of data and to find desired information, products and workers. The personalized recommender system, first proposed in the 1990s, is a powerful tool to alleviate the information overload problem by prioritizing the delivery of information and showing every user a different list of new items that match her personal interests and preferences (Lee and Brusilovsky, 2007). Recommender systems have been widely and successfully applied in online websites and e-commerce services. For instance, a customer on Amazon possibly sees a page called "Customers Who Bought This Item Also Bought," which displays the products that she is likely to be interested in. After people watched a movie on Netflix, it often suggests people what to watch later, called "People Who Liked This Movie Also Saw"(Jannach et al., 2010).[1]

Similar scenarios can be found on internet-based recruiting platforms, which have now accumulated a vast volume of information on workers and jobs. According to statistics from Glassdoor.com, in the US, there were 2.09 million jobs posted online by employers in 2019, and more than half of job seekers preferred finding job opportunities on online job sites.[2] In addition, the wide usage of online job searching and recruiting enables internet job boards to characterize behaviors and activities of job seekers and employers, which together foster the development of job recommender systems. Job recommender systems apply the concept of personalized recommendation to the job recruiting domain to suggest better matches between job seekers who search for job positions and recruiters who find candidates on the Internet. Virtually all internet job boards now recommend jobs to

[1] Recent evidence shows that 35% of purchase on Amazon and 80% of stream time on Netflix are driven by the recommendation systems. See `https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48` and `https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers`.

[2] See Glassdoor's HR and Recruiting Stats for 2020 `https://www.glassdoor.com/employers/resources/hr-and-recruiting-stats/reasons-to-use-glassdoor`, and Glassdoor's Job & Hiring Trends for 2020 `https://www.glassdoor.com/research/app/uploads/sites/2/2019/11/Job_Hiring_Trends_2020-FINAL-1-1.pdf`.

the workers who use their platforms. These customized recommendations are generated by algorithms, using criteria that include the worker's characteristics and previous behaviors, and the match between the worker's characteristics and the job's requirements. While job recommendation algorithms have the potential to help workers and firms find better matches faster, they also have sparked deep concerns about fairness: even when there is no discriminatory intent from designers, the recommended jobs may reinforce gender and other stereotypes. For instance, in content-based recommendation algorithms, gender might be associated with certain types of jobs and specific personalities in the workplace, which leads to gender segregation in job recommendations (Chaturvedi et al., 2021; Gaucher et al., 2011). Furthermore, based on job seekers' application behaviors, item-based collaborative filtering algorithms, as well as algorithms that incorporate the past behaviors of hiring agents, can create and perpetuate previous gender differences in recommendations received by workers.

This paper measures whether, to what extent, and how job board algorithms systematically treat male and female job seekers differently by conducting an *algorithm audit*, which is a new research approach proposed in recent years to study the black-box of algorithm features and to ascertain whether algorithms result in harmful discrimination by using fictitious correspondence in online platforms (Sandvig et al., 2014; Hannák et al., 2017). More specifically, I created otherwise identical male and female worker profiles on the four largest Chinese job boards, and observed which jobs were recommended to those profiles. In each job board, I selected 35 types of jobs based on three criteria: the number of active job openings, the job's gender-type (female-dominated jobs, gender-balanced jobs and male-dominated jobs), and hierarchy level (entry, middle, and high). Then I created resumes that were qualified for the above jobs; these come in pairs that are identical except for applicant gender. Since Chinese employers' gender preferences appear to interact strongly with the worker's age (Helleseter et al., 2020), I made two versions of each profile pair — a 'young' version and an 'older' version, in which the older applicants have 10 more years of working experience than young applicants. In order to track how algorithms update their recommendations based on workers' application behaviors, my fictitious workers then applied for the top jobs in their recommendation lists. I repeated this application process up to three times (each time responding to a new set of recommendations), then compared the job recommendations received by male and female applicants.

I find that identical male and female applicants do not always receive the same job recommendations: out of 100 job recommendations received by my applicants, 9.72 jobs were uniquely displayed to male or female applicants. Senior workers, who have more years of working experience, received a smaller number of gender-specific recommendations. Importantly, gender divisions in recommendations are even higher after fictitious applicants started applying for jobs: The raw difference rate between male and female applicants is 7.7% in the first round, whereas after three rounds of applications, 18.4 percent of recommendations are gender-specific. Because jobs displayed at the top of the recommendation list receive more attention, I further define the *list difference* in job recommendations, in which two job recommendations are the same only if both the job and the rank are identical in the recommended lists for pairwise workers (i.e., the third job in the men's list is the same with the third job in the women' list), and find that around three in four recommendations are different across male and female applicants.

To detect gender bias in the quality of recommended jobs, I leverage statistical tests to quantify the gender gap of both explicit and implicit measures of job quality. *Explicit measures* include the job's posted wage, requested education, and requested working experience. I find that on average, only-to-male jobs, which are seen by men rather than women, posted wages that were 1.9% higher than jobs recommended to women; this difference is marginally statistically significant. While the requested education is the same in jobs recommended to male and female applicants, jobs recommended only to men have 0.08 more years of working experience requirement than only-to-female jobs.

Furthermore, since job descriptions implicitly convey information on job quality, I extracted words used in the job descriptions reflecting five aspects pf quality: *skills*, *benefits*, *work form*, *company information*, and *other requirements*. By comparing the word frequency in male-only and female-only job ads, I find that literacy skills and administrative tasks are more likely to show up in female-only jobs, while influencing skills such as leadership and decision-making are mentioned more in male-only jobs. On the other hand, female applicants are recommended to apply for more jobs with flexible working hours and normal breaks in comparison to men with identical characteristics, while male applicants see more jobs that need night work and overtime. For benefits, only-to-female jobs place more emphasis on base pay, marriage leave, and parental leave, while only-to-male jobs focus on more performance incentives such as reward and company stocks or options. Company-related words do not significantly differ between male-only and female-only jobs, except that orientation training is involved in more female-only jobs, while male-only jobs are more likely to be in publicly-listed companies.

The other requirements contained in the job descriptions also reflect gender-based differences in job recommendations. Words in jobs recommended to women are often related to feminine personality, such as *patient* and *careful*, and have more descriptions on desired workers' appearance such as *facial features*, *figure*, and *temperament*. Jobs recommended to men prefer workers who are *self-motivated*, *experienced*, and are able to *work under pressure*. Moreover, these male and female words in recommended jobs are consistent with gendered words summarized in previous literature in language (Fitzpatrick et al., 1995), in political science (Roberts and Utych, 2020), in psychology (Rudman and Kilianski, 2000) and in labor economics (Gaucher et al., 2011; Kuhn et al., 2020; Chaturvedi et al., 2021). To collect the gendered perceptions of words, I conducted two surveys on Amazon MTurk and on Chinese workers, and found that feminine words emerge more within jobs seen by female applicants and jobs recommended to men contain more masculine words. This suggests that words used in gender-specific jobs are associated with widely held gender stereotypes in the workplace, and the inclusion of stereotype-linked words contributes to the gender bias in job recommendation systems.

Finally, I attempt to isolate the precise mechanisms accounting for gender bias in job recommendations. *Content-based recommendations*, which link gender with jobs' features must play a role because words about gender-related personality traits (i.e., patient in female, work under pressure in male) and gender stereotypes in the workplace (i.e., women are good at literacy skills, men have leadership) occur differently in gender-specific recommendations. Moreover, hiring agents' behaviors also appear to contribute to gender-biased job recommendations. When more hiring agents read their profiles, the pairwise male and female applicants will see more different job ads in their recommendations, indicating that human bias may be maintained in and interact with recommender systems. Lastly, by comparing jobs recommended before and after workers apply for jobs, I find that *item-based collaborative filtering* which recommends jobs based on workers' application history may reinforce and amplify the gender bias in the system.

This paper is related to four existing literatures. The first is the broad literature about gender inequality in labor markets. Using both traditional survey data and internet job board data, this literature has documented that gender inequality is accentuated by gender differentials in job search patterns, such that women are less likely to search for jobs outside of their living places and switch occupations (Eriksson and Lagerström, 2012), and women have higher levels of risk aversion in accepting offers (Cortés et al., 2021), from gender discrimination in the recruiting process in which employers prefer men in some certain occupations (Booth and Leigh, 2010; Cediey and Foroni, 2008), from gender segregation in skills (Christl and Köppl-Turyna, 2020; Stinebrickner et al., 2018), from gender differences in workplace bargaining propensity (Card et al., 2016), and from family burdens in promotions and career development (Petit, 2007). As far as I know, this is the first paper to study gender bias in job recommendations. While existing literature studies gender differentials at various stages of the search and matching process, I argue that gender differences and gender discrimination can arise even at the very early stage, where male and female workers may see different job vacancies in online job platforms due to the personalized job recommendations. More importantly, when the algorithm predicts workers' preferences based on their previous behaviors, feedback loops and self-fulfilling prophecies in recommendation algorithms may magnify the gender bias (Cowgill, 2018; Jiang et al., 2019), in which gender differences in job applications can yield to greater gender bias in the future job recommendations.

Methodologically, my paper contributes to the audit studies (or correspondence studies), which are widely used in the research on discrimination in social sciences. Aiming at comparing callback rates from real employers between two identities, audit studies have to create resumes that are as close as possible to real workers, and the detailed information, such as working experience on resume, is always randomly selected from resume banks (Gaddis, 2018). Due to the complexity of resume design and the high cost of callback collection, most audit studies only focus on a few occupations and industries, especially entry-level and unskilled jobs in manufacturing and service sectors; therefore, evidence on gender discrimination is lacking for senior-level and high skilled jobs which require proof of identity or qualifications (Rich, 2014). Compared to previous audit studies, my algorithm audit has three advantages: First, my resume design is much easier as the fictitious resumes only include the minimum information that is required by job platforms rather than any detailed descriptions of workers' personal working histories and statements. Second, since my goal is to investigate job recommendation outcomes from the workers' side, this study has no contacts with employers and does not collect callbacks from employers, which avoids alerting employers to the experiment (Avivi et al., 2021). Finally, the field experiment was performed on the four largest Chinese job boards and chose 35 job types in each platform, ranging from unskilled jobs such as sales and warehouse keeper, to high-level jobs such as financial manager and software engineer, which covered a broad and representative sample in online labor markets in China.

In addition, this paper contributes to the emerging literature on algorithmic fairness in economics. With the increasing engagement of algorithms in supporting human decision making, algorithmic bias and fairness have been studied in various fields such as advertisement delivery (Lambrecht and Tucker, 2019), criminal courts (Angwin et al., 2016) and mortgage approval (Fuster et al., 2020; Bartlett et al., 2021). In labor markets, existing research mainly focuses on gender bias in algorithms used in recruitment and in performance evaluation. For instance, Li et al. (2020) develop a resume screening algorithm that explicitly values exploration and show that efficiency (the quality of interview decisions) and equity (demographic diversity of applicants) can be improved at the same time in the workplace. Prassl (2018) documents that the evaluation algorithms in Uber result in lower payments for female drivers. However, to my knowledge, there is no research about the fairness of job recommendation algorithms from the perspective of job platforms. My research fills this gap by demonstrating that gender bias exists in the job recommendation algorithms, which comes even before workers apply for jobs. Moreover, when the hiring agents' behaviors are incorporated into the job recommender, gender discrimination in recruitment and gender bias in job recommendation interplays with each other, which potentially perpetuates gender inequality in the matching in labor markets.

Lastly, my work complements and extends research on gender equality (Poutanen and Kovalainen, 2017; Barzilay and Ben-David, 2016; Athreya, 2021; Cook et al., 2021) and algorithm transparency (Tambe et al., 2019; Kellogg et al., 2020) in the platform economy. From a practical point of view, few platforms in two-sided markets directly use information about gender, race or ethnicity in their algorithms. In other words, algorithmic bias is caused inadvertently in most cases. My empirical evidence from a strictly controlled experiment has important implications for platforms and policymakers to raise their awareness of the potential dangers of systematic bias in the algorithms.

The rest of this paper is organized as follows: Section 2 discusses the related literature. Section 3 provides an overview of how the job recommender systems generate job recommendations to job seekers. In section 4, I present the potential mechanisms of the gender-biased job recommendations in online job boards. Section 5 details the experiment design and implementation. Section 6 summarizes the experimental results on the differences in job recommendations between male and female applicants. I explore the potential drivers of algorithmic gender bias in job recommendations in section 7. Section 8 concludes.

2 Literature Review

2.1 Gender Discrimination and Audit Studies

Audit studies, also known as correspondence studies or correspondence experiments, have been widely used to estimate discrimination on various grounds, such as race, gender and age (Fix et al., 1993).[3] In recent audit studies on gender discrimination in labor markets, researchers create fictitious workers that are identical in all dimensions except for gender and send out their resumes to real job vacancies, then any difference between male and female job candidates on the subsequent callbacks from employers can be interpreted as causal evidence of gender bias or discrimination (Gaddis, 2018; Baert, 2018).

Empirical evidence on gender discrimination under the framework of audit study is mixed with respect to occupation, skill level, and age. An early work from Riach and Rich (2006) used pairs of matched, written applications to test for gender discrimination in London and showed that men had fewer callbacks in female occupations, and significant discrimination against females was found in male-dominated occupations. Similar results come from Booth and Leigh (2010), suggesting that the pro-female bias exists in the occupations where the percentage of females is 80% or more in Australia, Albert et al. (2011) documenting that females are significantly preferred in lower-level, female-dominated jobs in Madrid, and Carlsson (2011) showing that women have a larger advantage in female jobs than the advantage of male in male-dominated jobs in Sweden.

Moreover, Baert et al. (2017) show that when applying for jobs at a higher occupational level, the invitations for job interviews for female applicants are about two-thirds of that their male counterparts can receive in business-related jobs in Belgium. Using the three largest Chinese job websites, Zhou et al. (2013) describe the gender discrimination heterogeneity across firms: State-owned firms prefer male applicants due to leadership,

[3] More specifically, audit studies rely on real auditors who are matched in observable characteristics, while correspondence studies create and send fictitious applications with identical variables (Bertrand and Duflo, 2017).

while foreign firms, firms offering marketing positions, and short-lived private firms tend to interview more female applicants.

By conducting a correspondence study in France, Petit (2007) investigates the relations between family constraint and gender discrimination in hiring. It suggests a 20% gender gap in access to job interviews in which young, single female applicants (aged 25) are less favored in high skilled administrative positions, especially in jobs offering long-term contracts, but the discrimination is eliminated in prime-age applicants (aged 37) with children. In addition, being pregnant has a substantially negative effect on the probability of being interviewed in Belgium (Capéau et al., 2012), and mothers are penalized by a lower callback rate compared to childless women and fathers in the United States (Correll et al., 2007).

2.2 Gender and Internet Job Boards

While there is plenty of literature on gender differentials and gender discrimination in labor economics (Parsons, 1991; Keith and McWilliams, 1999), the expansion of online job platforms opens up new research topics and accumulates rich sources of data on job seekers and recruiters. With respect to the worker's side, internet job boards can follow the behaviors of job seekers through the whole searching process, which allows researchers to observe and compare the labor market participation behaviors of men and women. Although there is no conclusive evidence on the gender difference in job search intensity, research built on data from online job boards has documented that women are more selective and restrictive in their choice of search area (Eriksson and Lagerström, 2012), comply more to the minimum required experience, are less open to occupational moves (Banfi et al., 2019), and are less likely to search for long duration (Faberman and Kudlyak, 2019). Moreover, results from field experiments conducted on online job platforms demonstrate the gender difference in competition and job-entry choices. Flory et al. (2015) find that women are less likely to apply for jobs with competitive compensation structure and greater earnings uncertainty. Gee (2019) find that women are more likely to finish the job application when the number of received job applicants is shown in the corresponding job posting in LinkedIn.

On the employers' side, the recruiting process that is recorded by online job boards can be divided into two phases: the attraction phase and the selection phase (Färber et al. ,2003). Attraction phase mainly refers to job posting behaviors, in which employers specify job characteristics in job ads to attract qualified employees. Kuhn and Shen (2013) studied the gendered jobs in China, which explicitly listed the gender preference in job advertisements, to examine gender discrimination. They find that men and women are equally preferred in gendered jobs, but the preference for females to males often links to youth, height, and beauty rather than offered wages and skills. In the followed studies, Helleseter et al. (2020) documented the age twist in employers' gender requests, in which gender preference shifts away from women towards men as the target age of worker rises. When employers select suitable job candidates from applicants' pool, most of the employers make callbacks to applicants with requested gender, and the gender mismatch penalty is greater for women than men (Kuhn et al., 2020). In particular, after removing the gender label in job ads, the application rate and the success rate of jobs that requested opposite gender increases for both men and women (Kuhn and Shen, 2021).

2.3 Algorithmic Fairness

Algorithmic decision-making is increasingly engaged in social and economic life, and the question of algorithmic fairness attracts plenty of research from computer science and social science. For instance, the application of algorithm tools may lead to racial bias against black defendants (Angwin et al., 2016; Cowgill, 2018), racial/ethnic discrimination in mortgage, lending and credit approval (Bartlett et al., 2021; Fuster et al., 2020), racial discrimination in health system (Obermeyer et al., 2019), and gender disparity in image search and face recognition (Kay et al., 2015; Klare et al., 2012).

Interestingly, most of the unfairness is not intended by the algorithm designers. One of the main reasons for the bias is the input data, which can be biased or unrepresentative (Kim, 2017). If the algorithm is trained on data produced by biased human decision-makers, it will reflect the bias and probably deliver bias results, as the saying goes, Bias in, bias out (Rambachan and Roth, 2019). When the characteristics for some certain groups are missing or underrepresented in training data, the algorithm's prediction on these groups is likely to be inaccurate or biased (Barocas and Selbst, 2016). In addition, interactions between users, and interactions between users and platform can also contribute to the biased results(Jiang et al., 2016).

Recently, there is growing literature about the fairness of algorithms applied in hiring. One line focuses on the adoption of algorithmic decision tools in employee selection, such as resume screening, AI interviews, evaluation on interview performance, and productivity prediction (Mann and O'Neil, 2016; Lee and Baykal, 2017; Chalfin et al., 2016; Tambe et al., 2019; Li et al., 2020). However, the resume screening tool developed by Amazon was criticized for its higher ratings for male candidates than females, which resulted from the biased training data in which Amazon hired more male workers in the past.[4] Based on the investigation on 18 vendors of algorithmic pre-employment assessments (i.e., questions, video interview analysis, and gameplay), Raghavan et al. (2020) found that most of the vendors made abstract references to "bias", but few of them explicitly revealed how to validate their models and how to fix the bias in practice.

The other line is about employers' reliance on internet platforms, and the closest work to this paper comes from Lambrecht and Tucker (2019), who conducted a field experiment on Facebook to test how online advertising algorithm delivers STEM job opportunities differently to men and women. They ran advertising campaigns targeting both men and women with otherwise identical backgrounds and found that ad about job opportunities and training in STEM was shown to 20% more men than women. While the algorithm is intended to be gender-neutral, it creates gender-biased results as a consequence of optimization cost-effectiveness in ad delivery. In advertising auction, female eyeballs have

[4] Jeffrey Dastin, Amazon scraps secret AI recruiting tool that showed bias against women, `https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G`

more bidders and a price premium has to be paid to show ads to women relative to men, so the STEM ads were crowded out by other advertisers in the competition. Similar results are replicated by Ali et al. (2019), who show that ads can be delivered to vastly different racial and gender audiences when Facebook optimizes for clicks. For instance, with identical ad target options, jobs in the lumber industry were delivered to an audience that was 72% white and 90% male, jobs from taxi companies reached 75% Black users, and ads for cashier positions were shown to the audience of 85% female. While the two studies focus on gender inequality under a framework of price auction in which job opportunities compete with consumer goods in commercial advertisement delivery, I study gender bias in job recommendation algorithms on internet job boards, which are the dominant platforms that match workers to jobs. Another relevant work comes from Chen et al. (2018), who explore gender equality in ranking algorithms in resume search engines using data on 855K job candidates from Indeed, Monster, and CareerBuilder based on 35 job titles in 20 U.S. cities. They find that there is a slight penalty against feminine candidates even after controlling for all other visible candidate characteristics in resumes. On the group level, the unfairness significantly benefits men in 12 out of 35 job titles. In the setting where employers proactively search for workers, ranking algorithms affect job seekers' opportunities when employers contact more the top-ranked workers (clicked into their resumes), but my paper probes into the gender inequality problem in the aspect of job seekers' applications and argue that multiple channels, more than employers' behaviors, can contribute to the gender differences in job recommendations (in terms of recommended jobs as well as their ranks).

3 An Introduction to Job Recommender Systems

Before summarizing the job recommendation algorithms, I first describe the setting in which the job recommender systems work. On internet job platforms, when a job seeker with a complete profile logs into her account, the website displays a list of jobs that the job seeker may be interested in on her homepage. Unlike the search function that requires job seekers to input keywords in the search bar, job recommendation systems generate and present recommendation results proactively and automatically.[5] Based on my personal experience with several job boards and the academic surveys on job recommender systems from Al-Otaibi and Ykhlef (2012), Hong et al. (2013) and Siting et al. (2012), most of the online job platforms build hybrid recommender systems that incorporate multiple methods.

The core and foundation of most recommender systems is *item-based collaborative filtering* (item-based CF) method. Item-based CF uses the implicit collaborations of users or items to predict the users' preferences and filters the items that are most likely of interest to users; which can be expressed as "Users who liked this item also liked". The main idea of item-based CF is to recommend items to users that are similar to the ones that the users liked in the past, where the similarity between items is derived from users' rating behaviors (Jannach et al., 2016). In the context of job recommendations, users' rating behaviors reveal how the job seeker likes a certain job. Rating behaviors, such as clicking into a job, viewing the job page, marking the job as favorite, sending a message, indicate how the worker likes a job, and of course applying for a job indicates the job seeker's strongest preference for that job. If two jobs X, Y are applied (liked) by the same job seeker, they should share some features that attract workers to apply both of them, so the two jobs can be defined as "similar", in other words, two jobs are "similar" if they have enough overlapped applicants. To sum up, item-based CF recommends the jobs that are similar to ones that the target job seeker applied to in the past to the job seeker.

Pure item-based CF does not function well when user's behavior data is unavailable or very sparse (e.g. newly registered job seekers and newly posted jobs).[6] To deal with

[5] Job recommendations play an increasingly important role in online job boards because Boolean search methods only adopt keywords to generate the results, which is insufficient and may fail to generate an appropriate match for workers and jobs (Lang et al., 2011).

[6] If no user information is available, for instance, browsing as guests or newly registered job seekers, the knowledge-based recommendation will be used to list jobs that satisfy the user's requirements on jobs, such as the job's location, wage, and occupation. For newly posted jobs, it uses two methods to overcome the ramp-up problem. One is to apply content-based recommendations to find old jobs that are similar to the new one and recommends the new job to applicants who have already applied to these old jobs. The second is to rely on cooperation with the search algorithm. The search algorithm gives more weights for newly posted jobs to encourage job seekers to apply for new jobs when they contain certain keywords of the

the cold-start problem, most job recommender systems use a *content-based* algorithm as an important supplement to item-based CF. Based on text analysis and natural language processing techniques, content-based recommendation algorithms identify similarity between two documents by comparing the keywords in the documents, in which "content" refers to the descriptions of items' characteristics and the users' profiles.[7] In online job boards, content similarity can be established between jobs, between workers, and between jobs and workers. Two jobs are defined as similar when the same keywords appear in the job descriptions. If a job seeker applies to one of the jobs, similar ones will be recommended to him because he should have a consistent preference on jobs' content. Two workers are similar if their resumes have the same keywords, and the jobs one job seeker applies to will be recommended to the other one, since two job seekers with similar resumes should share similar tastes. Moreover, the content-based method also utilizes *job-worker match* attributes to make recommendations. For example, if a job ad and the worker's resume contain the same keyword, such as a skill, then the system will suggest the job seeker to apply for that job.

A third method used in job recommender systems applies a *rule-based approach* to the rich information on jobs and workers on online job platforms to make recommendations based on the match between jobs and workers. The rule-based approach frames job recommendation as a classification problem and relies on worker's characteristics and job's requirements to predict the fit between the target worker and a certain job. For instance, if a worker satisfies the education requirement of a job, the job website is more likely to recommend this job to the worker.

Finally, some job boards apply more sophisticated systems that incorporate the hiring agents' behaviors into recommender systems and suggest jobs the target worker is likely to get feedback from (Kim, 2017). From the perspective of job boards, job seekers may

request of job seekers. After gathering some initial ratings, these new jobs will enter into item-based CF and can be recommended to other job seekers.

[7] Some researchers frame content-based recommendations as a classification problem of the user's likes and dislikes, and the goal is to find the classifier based on item characteristics. In this line, lots of supervised machine learning techniques such as Bayesian Classifiers, clustering, decision trees, and artificial neural networks can be applied to train models which can automatically decide whether a user is going to like a certain item.

become frustrated by sending out lots of applications but getting no echo, and switch to other sites as a result. Therefore, these *recruiter-behavior based algorithms* use recruiters' rating information to determine which type of jobs require which type of workers' characteristics and the probability of the worker getting callbacks when making job recommendations (Al-Otaibi and Ykhlef, 2012). More specifically, platforms collect the recruiter's application processing behaviors and predict the recruiter's preferences based on those behaviors. If the recruiter produces some positive signals towards a certain job applicant, such as browsing or downloading her resume, the system will acknowledge that the job prefers that type of job candidates, and recommend this job to workers who are similar to that job applicant (Yu et al., 2011). Moreover, the worker will receive job recommendations that are similar to this job since she has a relatively high chance to be suitable in similar positions.

4 Potential Mechanisms for Gender Bias

Although the algorithms used in job recommender systems are theoretically intended to be gender-neutral, there are at least four ways that recommender systems can deliver gender-biased job recommendations, which are connected to the four main components of most current job recommender systems.

The first is from item-based collaborative filtering recommendation, which recommends jobs that are similar to ones that the worker applied to in the past. While not in itself gender-biased, this algorithm tends to magnify and perpetuate previous gender differences in recommendations received by the worker. Suppose there is a job requesting male workers. In the extreme case, due to the gender mismatch, the job is not recommended to any female workers, and no female workers can see and apply for the job. In the following job recommendations, the absence of this job in female workers' application histories will reduce the exposure of other jobs that are similar to that job, even without gender request, and induce more divergence on the recommendation results between two genders.

The second component is content-based recommendations among workers. It is worth noting that the foundation of content-based recommendations, natural language processing algorithms can embody gender bias. For instance, female names are more associated with family than career words, compared with male names (Nosek et al., 2002). Nurse, teacher are more likely to be associated with she or her, while engineer, scientist are associated with he or him, suggesting that implicit gender-occupation biases are linked to gender gaps in occupational participation (Caliskan et al., 2017). If this is the case in job boards, we may observe some jobs are recommended to one gender more frequently than to the other gender because their characteristics are encoded to be correlated with gender identity, and the algorithm eliminates workers whose resumes do not contain the gender-related keywords (Savage and Bales, 2016). Furthermore, if the keywords associated with strong gender tendency are used to define similarity between workers, workers with the same gender consequently are more likely to be similar. For instance, patient is found in the resumes of female workers more often (or expatriate in male resumes), and if the algorithm uses these kinds of characteristics as the keywords in contents, workers are classified based on gender (Bozdag, 2013). As a result, female workers may be recommended with jobs that have been applied by other females, leading to gender segregation in job recommendations. Importantly, when jobs having gendered keywords are defined as similar, a worker that applies for one job with gendered words, will be recommended to other jobs that also contain such gendered words.

The third mechanism relates to the rule-based approach, which frames job recommendation as a filtering problem and only considers the 'hard' match between the worker and the job. If a job's characteristics are consistent with the worker's expectations and the worker satisfies the job's requirements, the job will be recommended to that worker. One important feature of Chinese job platforms is that they allow employers to explicitly state the gender of preferred applicants, without revealing these preferences to job seekers in the ads. Thus, for example, a rules-based algorithm might not show ads that list a preference for women to male job seekers.

Finally, consider recruiter-behavior based approach that incorporates hiring agents' rating behaviors (i.e., viewing and downloading profile, sending a message to target worker) into the recommender system. As far as I know, there are three scenarios in which the hiring agents' behaviors could affect job recommendations.[8] Suppose a hiring agent posted a job and received some applications from both genders, but has consistently ignored female applicants (for example, never downloaded female resumes).[9] Two points are learnt from this process: First, this job is not going to hire female workers, then it will not be recommended to other female workers. Second, if a female applicant did not get positive feedback from the job, the algorithm infers that she is unlikely to get callbacks from other jobs that are similar to that job, so those similar jobs will not be recommended to her. That is to say, workers' recommendation results are affected by the processing decisions of the hiring agents who posted jobs that they have already applied to, as well as the spillover effects from other hiring agents. Moreover, in most online job boards, hiring agents can search for and contact suitable workers directly. When a hiring agent searches for workers and clicks into a worker's resume, the jobs posted by this hiring agent will be recommended to that worker, as the hiring agent has shown interests to that worker (Köchling and Wehner, 2020). If a hiring agent persistently views resumes of male workers, those male workers will be suggested to apply while female workers do not have this priority (Burke et al., 2018).

The mechanisms mentioned above can interact with each other to create a complex job recommendation system.[10] More generally, algorithms may replicate the errors stemming from the training data, such as choosing parameters based on data with existing stereotypes, which detracts from gender fairness. Overall, recommender systems may reproduce and magnify pre-existing gender bias in the labor market.

[8] Algorithms targeting at click maximization are likely to deliver biased results, due to the feedback loop (Jiang et al., 2019) and learning-to-rank approach (Jiang et al., 2016; de Sá et al., 2016).

[9] The four online job boards allow recruiters to filter workers' profiles by demographics (e.g., gender, age) and characteristics (e.g. education, experience) when they process received applications or search for suitable candidates.

[10] Both direct discrimination and indirect discrimination on gender potentially exist in these algorithms, which are distinguished by whether sensitive features (gender) are not explicitly used as inputs in algorithms (Pedreshi et al., 2008).

5 Experiment Design

5.1 Platform Environments

To cover a representative sample in online labor markets, the experiment was conducted on the top four job boards in China, which have millions of job seekers and job postings and can reach most of the workers and recruiters in the Chinese labor market. The large consumer bases allow me to create substantial fictitious workers but minimize the disturbance of the job search and recruiting process as well as the job recommender systems. The four job sites have similar interfaces and functions for users, with regular structures of online job platforms. Job seekers can register and create a profile for free, while employers are charged for posting job advertisements and using recruiter tools. Job seekers make applications by sending their resumes to the jobs that they are interested in, and hiring agents of firms can check and process the applications online and contact applicants through the website's message system. Furthermore, as far as I know, the leading job boards use more detailed and sophisticated forms of machine learning to suggest jobs to workers, and I may expect that the advanced algorithms may reinforce gender bias in an implicit way.

5.2 Job Type Selection

When a job seeker sets up her profile, job platforms let her indicate her current and desired industry and occupation. This job type information will be used by the job recommender systems and affect job recommendation results.

The selection of job types is based on three criteria: sample size, gender type, and hierarchy level. As a first step, I chose industry-occupation cells that have a large number of job postings to ensure that there were enough new job vacancies to be recommended to workers.[11] For instance, the internet industry has the most job postings, while sales are the

[11] The industry-occupation cell refers to the sub-industry and sub-occupation because workers will choose

most popular occupations in job sites, so the internet-sale is a potential job type. Second, because male-and female-dominated jobs might prefer applicants of different genders, I focused specifically on three gender types of jobs: female-dominated (i.e. administrative assistant), (approximately) gender-balanced (i.e. sales), and male-dominated jobs (i.e. software engineer).[12] Finally, because employers' gender preferences may also vary across the job ladder in which few women reach the top positions on the job ladder (Bertrand et al., 2010; Pekkarinen and Vartiainen, 2006), I diversify the hierarchy by including jobs in entry-level, middle-level and high-level. Taking the job of sales as an example, salesclerk is the entry-level job, sales manager is a middle-level job, and sales director is a high-level job. The details of these job types and the related characteristics of workers are described in Appendix A1.

5.3 Resume Setup

I next created resumes that are qualified for the above jobs. The fictitious resumes come in pairs, and the two workers in each pair have identical backgrounds, except that one is female and the other is male. These resumes are quite sparse and contain only the mandatory information that is required to set up a worker profile to make sure that the recommendation results are not driven by other details. To achieve valid job recommendations, resume information is generated based on the real job ads and workers' resumes. For each job type, I scraped 50 job ads and 50 resumes as the information pool for fictitious profiles.

A fictitious applicant's resume consists of four parts: personal information, education, job history, and job intention. Personal information section collects worker's name, birth date, years of working experience, current wage, city, employment status, phone number, and email address. Different from most audit studies relying on workers' names to signify gender identity, gender (male or female) is a compulsory input in Chinese job

the finest category of industry and occupation when they set up their profiles.

[12] The selection of job's gender type is based on the public statistics and reports on the share of female workers in job boards.

boards. The applicant's name is randomly assigned with the most popular names from 2015 Chinese Census 1% Population Sample, and the first name is matched with gender (See Appendix A2.1 for more details). Since Chinese employers' gender preferences appear to interact strongly with the worker's age (Helleseter et al., 2020), I create two versions of each matched profile pair—a 'young' and an 'older' version, in which 'older' workers refer to ones who have more working experience. Worker's age, education and working experience are jointly determined. Young workers graduated in 2017, have three years of working experience, and are either 25 years old (born in 1995) if he has a college degree, which takes three years to achieve, or 26 years old (born in 1994) if he has a bachelor's degree, which takes four years to achieve. The corresponding older workers are 35 (college) or 36 (bachelor's) years old with 13 years of working experience. The specific education level and academic major satisfy the requirements of job type, and the school's name is randomly drawn from the Chinese High Education Institution List.[13] All the applicants are currently employed, and their wages are crafted to match the wages of existing job seekers by job type, education level, and years of working experience. As over half of job postings are from first-tier cities, I restrict the location of applicants to the first-tier cities in China, including Beijing, Shanghai, Shenzhen, and Guangzhou. Each applicant has a unique and active email address and mobile phone number.

In terms of job history, young workers started their current jobs in August 2017, just after they graduated with the highest degree. For older workers, the beginning date of their current jobs is August 2015, implying that they have 5 years tenure in their recent positions. Worker's current occupation and industry are the same as the job type's occupation and industry, and job title and job description are entered as the job's occupation. I make up the company name to minimize the disturbance to both job seekers and employers in job platforms, which is a combination of worker's city, industry and a randomly generated name (i.e., Beijing Dongya Internet Technology Company). In the job intention section, a worker's desired wage is 120% of his current wage, and the desired city, industry and occupation is aligned with current ones.[14] Appendix A2 summarizes the details

[13] Released by the Chinese Ministry of Education in 2019.

[14] According to the salary reports from the job boards, 20% is normal and moderate wage growth for an

of resume generation process.

To sum up, I created groups of four resumes that vary along two dimensions, gender and age, with all the other characteristics and information held constant, except that the older resumes' experience and current wages are adjusted to be age-appropriate. Given that the four workers in each group are designed to have the same job type and 35 job types are selected in each job board, I created 140 fictitious profiles (replicated across 4 cities) on each platform. After finishing the profiles creating process, male and female applicant published their profiles at the same time, afterwards their resumes are accessible (can be read or downloaded) to recruiters and headhunters on the platforms.

5.4 Implementation

In addition to workers' resume characteristics, job recommender systems user workers' browsing and application behaviors to deliver customized recommendations. To control for such differences, the paired (male and female) profiles followed identical application strategies. Fictitious workers are naïve users on the job platforms, who click into and send resumes to the top listed recommended jobs. This process works as follows:

- **Round 0**. The male and female workers with newly created resumes log into their accounts at the same time, and I collect the first advertisement listed in the recommendation interface, to a maximum of 100 jobs. Then the workers log off.
- **Round 1**. The male and female workers simultaneously log into their accounts again, and I record the top 10 jobs (1st to 10th of listed job ads) in their recommendation lists. The two workers then apply to the top 10 recommendations by submitting their resumes. Immediately afterwards, the workers refresh their webpages and I record the 10 recommended jobs that appear.
- **Round 2**. At two-week intervals, I repeat the Round 1 procedures.

average worker switching to a new job.

- **Round 3**. After two weeks, I repeat the Round 1 procedures again.

- **Round 4**. After two weeks, male and female workers log into account at the same time, and I record the number of views on the worker's resume by hiring agents. [15]

Figure 1 demonstrates the timeline of this experiment. Ideally, each fictitious worker received 160 recommended jobs and applied for 30 jobs in an 8-week job searching spell, and the collected outcomes include the information of 160 jobs as well as the number of hiring agents' views on profiles. The design of my field experiment guarantees that any observed differences in the job recommendations are caused solely by my randomized gender manipulation.

6 Results

My audit study of job recommendation algorithms started in July 2020 and the last round of collections on hiring agents' views was completed in April 2021. In total, 2,240 fictitious profiles were created in four job sites, and those workers received 319,974 job recommendations from 119,356 individual job advertisements.[16]

Table 1 presents the descriptive statistics of my sample of fictitious workers. As applicants are designed in pairs and have fixed characteristics, Table 1 mainly reflects the presence of job boards in labor markets. The average annual wage of worker sample is 142,507 RMB, which is around twice the national average wage of workers in the urban

[15] After a worker's resume opens to the public, it can be viewed by all recruiters on the job boards. Recruiters of the applied jobs can read applicants' profiles, and other recruiters can find workers by searching resume, or by worker recommendations from job boards. The number of views records how many times that the resume is read by hiring agents.

[16] There are several reasons that the recorded number of job recommendations is smaller than the designed number 2,240*160 = 358,400. The first reason is in Round 0, some job types did not have 100 active job openings that matched the characteristics of workers. The second reason is that, job boards froze suspicious workers' accounts and a few of them were blocked after Round 0. If one account in a gender pair was blocked, I terminated the experiment of the whole gender pair. Another reason is job ads were withdrawn by hiring agents when I scraped job ads so some jobs' information was unavailable. The missing data is less than 0.5% and occurs randomly, especially it is independent of the gender of fictitious applicants, thus unlikely to bias my analysis.

in 2020.[17] The desired wage is 26.1% higher than the current wages,[18] and the average years of education is 15.56, indicating about half of the fictitious workers hold a bachelor's degree.[19]

The sample of recommended job ads is summarized in Table 2. Conditional on information is visible to workers, the average job posted an annual wage of 211,004 RMB, requested workers with 14.4 years of education and 2.3 years of working experience. On average, employers advertised a wage that was 17.4 percent higher than the fictitious workers' desired wages, but requested lower education levels and fewer years of working experience. In addition, above 95% of job ads have explicit wage postings,[20] and one-third of the recommended positions are from companies that have more than 1,000 employees. The fictitious workers' profiles are well-matched with recommended jobs, as shown in Table 3. More than 80% of designed workers satisfied the jobs' requirements on education and working experience, and almost all of the recommended jobs' locations aligned with the worker's current location. 86.6% of recommended jobs posted wages that were higher than workers' lowest desired wages.

6.1 Set and List Differences between the Job Recommended to Men and Women

This section answers the most basic question about gender bias in job recommendations: To what extent are the jobs recommended to male and female workers the same, or different? I quantify the gender difference in job recommendations in two dimensions: the set difference and list difference.

[17] According to the statistics from National Bureau of Statistics of China, the average annual wage of workers in the urban non-private sector in 2020 was 97,379 yuan (US$15,188), and workers in the urban private sector had an annual wage of 57,727 yuan (US$9,004).

[18] Some job boards let the worker choose desired wage range, and the desired wage is the midpoint of selected desired wage range.

[19] It is common to take 16 years to achieve a Bachelor's degree, and 15 years to achieve a college degree.

[20] While some empirical evidence suggests that better jobs (i.e. higher requirements on education and experience) are less likely to explicitly post wages (Marinescu and Wolthoff, 2020), it is not true in my data.

6.1.1 The Set Difference

Set difference measures the share of jobs that are only recommended to one gender, without considering the sequence of recommended jobs. Figure 2(a) demonstrates the set difference: Suppose for all workers, male applicants receive jobs that are in set A and C, and the female applicants are recommended by jobs in set B and C, in which set C contains the overlapped jobs of female and male recommendations, while set A represents the only-to-male jobs, and set B includes the only-to-female jobs. Then the set difference rate is defined as the share of only-to-one gender jobs on the whole pool of recommended jobs received by male and female applicants:

$$Set\ Difference\ Rate = \frac{\#\ jobs\ in\ A + \#\ jobs\ in\ B}{\#\ jobs\ in\ A + B + C} \tag{1}$$

I present the set difference rate by worker's age level, by job's gender type, by job's skill level and by city in Table 4. In total, the set difference rate between male and female applicants is 9.72%, meaning that out of 100 jobs recommended to male and female applicants, 90.28 jobs are displayed to all applicants, and 9.72 jobs are unique to one gender while applicants with the opposite gender cannot see those jobs in their recommendation lists.

While we expected that the gender difference in job recommendations would be greater in jobs typically occupied by males or females, our empirical results do not support that claim. In contrast, male and female applicants working in gender-neutral jobs observe about 1 additional different job per 100 recommended jobs, compared to workers in male- or female-dominated job types. For the age variation, young applicants who have 3 years of working experience are more likely to be exposed to gender-specific job ads, but the difference is quite small. Job hierarchy also matters job recommendations to men and women, in which gender-specific jobs appear more frequently in middle-level jobs, and least in entry-level jobs. The last panel of Table 4 shows the geographical evidence on the share of gender-specific jobs. Generally, the set difference rates are close across the four cities, indicating no spatial disparity on the gender difference in job recommendations is detected.

To confront the issue that the pattern of gender bias may vary across subgroups, I decompose the number of different jobs between two genders by age, job's gender type and hierarchy in Figure 3. Two features can be identified: First, Figure 3(a) illustrates that the gender difference in recommended jobs is greater for young pairs and more pronounced in gender-neutral jobs in older pairs. Second, in Figure 3(b), female-dominated with middle- and high-level jobs contribute a large share of different jobs between male and female applicants. Figure 4 further displays the dynamics of set difference rate. Without making any job applications, the share of gender-different jobs in Round 0 is 7.66 percent, and the share goes up after Round 1 when workers started to make job applications and is perpetuated with the application process. At Round 3, the chance of applicants viewing a gender-specific job is more than doubled relative to the share in Round 0.

6.1.2 The List Difference

While the set difference reveals the number of recommended jobs that are unique to one gender, it only partially uncovers the difference on job recommendations, because job ads are ranked in the recommendation list; ones displayed at the top receive more attention, and are more likely to be seen and clicked into by workers (Craswell et al., 2008; Richardson et al., 2007). For instance, if the jobs received by male and female workers are completely the same, but male workers observe the jobs ranked from high to low (in quality) and female workers see a list of an opposite order, we can hardly say that they achieve the "same" job recommendations.

Now I take the rank of recommended jobs into account to measure the gender inequality in job recommendations. Define two job recommendation lists are the same only if the two jobs in the same rank are identical, as shown in Figure 2(b). Then the list difference rate is defined as:

$$List\ Difference\ Rate = \frac{\sum_{i=1}^{n} ith\ job\ ad\ is\ different\ in\ gender\ pair}{Length\ of\ recommendation\ list\ (n)} \tag{2}$$

Table C1 summarizes the average list difference rates by experimental rounds, worker's age level, job's gender type, hierarchy level and city. The difference rate inflates after considering the ranks of jobs. The overall list difference rate is 70.7%, indicating that in a list of 100 recommended jobs, there are about 30 jobs that are displayed identically to male and female applicants. Similar to Figure 4, the list difference rate largely increases after applicants send out job applications, from 58.3% in Round 0 to 86.4% in Round 3. The list difference rate has a quite consistent pattern with set difference rate across the subsamples by age, job's gender type, hierarchy level and city.

6.2 Differences in the Quality of Jobs Recommended to Men and Women

As shown above, job recommendations to male and female workers are not the same. This dissimilarity does not necessarily indicate bias, however, because it could result from randomness in each website's recommender system. However, if systematic gender bias actually exists in job recommendations, jobs recommended to one gender would be better than jobs shown to the other gender. To address this question, this section explores whether job recommendations to the two genders are equally good.

6.2.1 Explicit Measures: Wage, Education and Experience Requirements

While jobs can be evaluated from various dimensions (Brenčič, 2012), I start from the explicit characteristics: the job's posted wage, requested years of education, and requested years of working experience. In order to compare the quality of different jobs recommended to male and female applicants, the subsequent analysis sample is composed of jobs unique to male applicants (i.e. jobs in Set A in Figure 2(a)) and jobs unique to female applicants (jobs in Set B), and the overlapped jobs (i.e. jobs in set C) are excluded.

I use the two-sample t-test to examine whether the mean of the characteristic in male-only jobs equals the mean in female-only jobs. Suppose in job ads received by all men and women, the observed job's characteristic x in male-only job sample is $(x_1^M, ..., x_n^M)$, and in female-only job sample is $(x_1^F, ..., x_n^F)$, where n denotes the number of different jobs in male or female job recommendations,[21] then the null hypothesis of two-sample t-test is:

$$H_0 : \overline{x^M} = \overline{x^F}$$

Taking the posted wage as an example, under the null hypothesis, the average posted wage of only-to-male jobs does not differ from the average posted wage of only-to-female jobs.

Table 5 presents the results of two-sample t-test on the job's posted wage, education requirement and working experience requirement.[22] Conditional on the wage is advertised publicly, the gender gap of recommended wage between male and female applicants is 2,709 RMB and significant at 10% level, which is equivalent to 1.9% of the average current wage of fictitious workers, meaning that jobs recommended to men propose higher wage on average the jobs recommended to women.[23] The requested education is statistically indistinguishable between male-only and female-only jobs, but the required working experience in male-only jobs is significantly higher than the requirement in women-only jobs by 0.08 years, which is translated into 0.5% of the average worker's working experience.

To facilitate comparison on the gender gap for subgroups, I provide the two-sample t-test results by experimental rounds, worker's age, job's gender type and hierarchy level in Figure 5. According to Figure 5(a), the jobs' wages in male-only and female-only recommendations do not differ among young workers and older workers, across female-

[21] n can be different for male and female applicants due to the replicated recommendations.

[22] Equal variance is applied, and results for two-sample t-test with unequal variance are in Appendix Table B1.

[23] Instead of an exact wage, most jobs posted a wage range. The job's wage in the analysis is the midpoint of posted wage range.

dominated, gender-neutral and male-dominated jobs, and across job levels, suggesting that gender effect on recommended wage does not interact with age and job type (gender and level). But regarding the rounds, in rounds 3, after workers applied 20 jobs, only-to-male jobs post significantly higher wages than only-to-female jobs. Similar to the previous results, the differences in education requests of recommended jobs to males and females are positive but remain insignificant in all subsamples, as shown in Figure 5(b). In Figure 5(c), the higher requirement on working experience in male-only jobs is pronounced in older applicants, in one gender-dominated jobs, and in entry- and high-level jobs.

Two-sample t-test assumes the variables are continuous and normally distributed or large sample size, and those assumptions might be violated in the analysis sample (i.e., the requested experience is ab integer). I provide the Wilcoxon rank-sum test as a robustness check in Appendix Table B1, which is a non-parametric test without assuming the certain distribution of variables, and the main results do not alter under Wilcoxon rank-sum test.

Turning to the list difference, I construct the comparison between male and female job recommendations by using paired t-test. After excluding the identical recommendations (i.e., 1st and 2nd job recommendations in Figure 2(b)), a paired applicants' job recommendation lists can be expressed as $\left((Y_1^M, Y_1^F), ..., (Y_s^M, Y_s^F)\right)$, in which s denotes the different recommendations. Suppose the ith job recommendation is (Y_i^M, Y_i^F), and Y_i^M and Y_i^F represents the ith job recommended to male and female applicants in the list-different recommendation sample. Define the difference d_i of job's characteristic y in ith recommendation as:

$$d_i = y_i^M - y_i^F \tag{3}$$

Under the paired t-test, the equally good job recommendations mean that the average difference between the characteristics in the two jobs listed in the same position is not significant from zero. The null hypothesis is:

$$H_0 : d_i = 0$$

Compared to two-sample t-test, paired t-test assumes that the two jobs recommended to male and female applicants in the same rank are correlated. Table C2 replicates the computation in Table 5 but replaces two-sample t-test with paired sample t-test. On average, jobs recommended to male and female workers in the same rank do not differ in their posted wages, required education and working experience.[24]

6.2.2 Implicit Measures: Words

In addition to explicit measures, a job's quality can be measured using the words in the job descriptions. Moreover, the presence or absence of a specific word may affect the matching between workers and jobs and leads to gender segregation in job recommendations (Dreisbach et al., 2019). In this section, I explore the gender difference in the wording in recommended jobs.

In a typical job ad, the job description is one or two paragraphs of text, which is placed after the explicit characteristics of jobs and contains rich information about the position. While the contents of job descriptions in different job types are highly diverse, they can be broadly aggregated into five categories:

(1) *Skills.* Skills are the core part of the job description, and recruiters express skills in various ways. For example, skill is often stated as a job requirement, "the candidate should be familiar with Excel", or a part of the position description, "common tasks include making reports with Excel". While plenty of methods are developed to deal with the complexity of skills in jobs, I adopt the skill classification in OECD Programme for the International Assessment of Adult Competencies (PIAAC) (OECD, 2016), which is widely used in the social sciences research on gender differentials on skills (Christl and Köppl-Turyna, 2020; Pető and Reizer, 2021). More specifically, skills are divided into seven subsets, including literacy skills, numeracy skills, information and communication technology (ICT), problem-solving skills, influencing skills, co-operative skills and self-organising skills.

(2) *Benefits.* In addition to offered wage, employers advertise jobs' benefits to attract ap-

[24] Similar to two-sample t-test, paired t-test requires that the measured differences are continuous and normally distributed. A robustness check with Wilcoxon signed rank test is presented in Appendix C2.

plicants. In Chinese job boards, the advertised benefits are often tagged, and their expressions are quite uniform across job types and platforms. Based on the extracted information from job ads, I classify job benefits into four types: payment, break, facility and insurance.

(3) *Work form.* Work form is a wide category that introduces the working time arrangement, capturing words about work schedule, business travel, work break and work overtime.

(4) *Company.* Job ads provide information on both the position and the company. The company features are summarized into three parts: workplace environment, company type, and title.

(5) *Other requirements.* Instead of simply displaying education level and years of working experience, employers state more detailed requirements in the job description, for example, the prospective candidates study in certain academic majors, have overseas working experience or have certain personalities. The other requirements category refers to the aspects of desired worker's age, appearance, personality, education, working experience, and other conditions.

Based on the above structure, the information in job descriptions was extracted in the following way: For all the jobs collected from four job boards, I first segmented a chunk of text into words (phrases) and retained words (phrases) with high frequency. Then I combined the words (phrases) that have the same or close meaning (i.e., leadership vs leading) to make the selected words (phrases) clearly contrast with each other, and assigned them to one of the five categories. In total, I extracted 167 individual words from job ads, listed in Appendix Table D1.

Figure 6 presents the word cloud of job descriptions, with the bigger size representing a higher frequency of words in job ads.[25] Words related to job benefits, such as insurance, vacation and payment scheme, are most commonly seen in job descriptions, and employers often state their requests on worker's communication skills, coordination skills,

[25] Appendix Figure D1 shows the word cloud in Chinese.

teamwork skills and leadership.

When the job recommendation is gender-neutral, the proportions of containing a certain word or phrase in male-only and female-only jobs should be close to each other, whereas when gender bias exists, some words will be differentially present in advertisements for jobs recommended to men and women. This hypothesis is examined by the proportion test: Word z is constructed as a binary variable, and z_i takes the value of 1 if job i contains the word z in its description. Under the proportion test, the null hypothesis is that the probability of the word showing up in male-only jobs equals the probability that it appears in female-only jobs.

$$H_0 : \overline{z^M} = \overline{z^F}$$

Table 6 displays the proportion test for wording difference in job recommendations under meaningful categories. The coefficient in parentheses represents the gender gap in words frequency ($\overline{z}^M - \overline{z}^F$), and a positive difference means that jobs recommended to male workers are more likely to contain that word in their job descriptions than female workers' jobs. The left panel lists 28 female words, which have a higher probability of being included in female-only jobs at 5% significance level, and the right panel includes 31 male words that are significantly mentioned more in male-only jobs.

We can see that most literacy skills, such as *speak* and *documentation* are more common in only-to-female jobs. Furthermore, female applicants are recommended for more jobs mentioning *data*, *chat tools* and *administrative tasks*, while male applicants see more jobs that require *problem-solving skills*, such as *decision-making*, *engineering*, and *working independently*, and *influencing skills* such as *leadership* and *manage*. These findings coincide with the results from previous literature on the gender gap in skills that document women tend to carry out more executed tasks, less skill-intensive tasks and use their cognitive skills less than men (Pető and Reizer, 2021; Black and Spitz-Oener, 2010).

In the *work form* panel, jobs with *regular working hours* or *flexible* schedules are more likely to be recommended to women, and jobs with decreased flexibility, such as *overtime working*, *night work* and *commute*, are more likely to be recommended to men. This is in line with the finding that women are more willing to pay for flexible work arrangements (Flory et al., 2015; He et al., 2021; Mas and Pallais, 2017; Bustelo et al., 2020). For *benefits*, female-only jobs are more likely to mention *base pay*, *marriage leave*, *maternity leave*, *social security*, *unemployment insurance* and *parental leave* in their descriptions while only-to-male jobs emphasize providing *shuttle*, *medical insurance*, *vacation*, *free meal*, *reward* and *stock*. Orientation *training* is mentioned more in female-only jobs, while jobs from the *publicly-listed* companies are more often to be recommended to male applicants. In addition to skills requirements, jobs recommended to men request workers who are *self-motivated*, *innovative*, *experienced*, and able to handle work *pressure*. Words associated with physical appearance, such as *figure*, *facial*, and *temperament*, and words about the feminine personalities such as *careful*, *patient*, *punctual*, *outgoing* and *trustworthy*, are more frequently emerge within female-only job advertisements. Jobs recommended to women also open to hire *new graduates* and workers *without working experience*, and request that the qualified applicants should be *healthy* and *below 35* years old. [26]

6.3 Words and Gender Stereotypes

The different presence of words in gender-specific jobs established above provides initial evidence that gender bias in wording exists in job recommendations. Previous research has shown that the wording in job advertisements reveals employers' preference on gender and would direct workers' application behaviors, even when employers do not make explicit gender requests. For instance, women found jobs less appealing when the job advertisements included more masculine wording (Gaucher et al., 2011), and feminine wording in job titles and job descriptions increases the share of female applicants (Kuhn et al., 2020; Chaturvedi et al., 2021). Job platforms may mediate employers' gender pref-

[26] The list difference in word frequency between male and female job recommendations was explored through McNemar's Test, and differences of the word usage between jobs recommend to only males and only females are insignificant in most cases.

erences in a special way in which the recommender systems target the desired workers by linking gender to the words in job descriptions, which can reinforce gender stereotypes and result in gender-based job recommendations. In this section, I further explore the relationship between gender stereotypes and words used in gender-specific jobs.

I rely on multiple external sources of data to identify the femaleness and maleness of words in job ads. The first source is the previous literature on gendered words, which refer to masculine and feminine words that are associated with gender stereotypes. While linguists focus on the commonly used words in daily life and the effect of gendered words on people's behaviors (Fitzpatrick et al., 1995; Gastil, 1990; Lindqvist et al., 2019), researchers in political science (Roberts and Utych, 2020) and in psychology (Bem, 1981; Hoffman and Hurst, 1990; Rudman and Kilianski, 2000) specify gendered words in various application scenarios and argue that the usage of gendered words would shape people's attitude and support to social values. Most of the gendered words identified by these literatures are adjectives, which describe men's and women's personalities (e.g., masculine words: confident, aggressive, strong vs feminine words: sensitive, kind, beautiful). More relevant to the current context, three papers have encoded the gendered words used in job advertisements and demonstrated the subsequent labor market outcomes. Gaucher et al. (2011) collected masculine and feminine words from published lists of agentic and communal words, and masculine and feminine trait words. Given the existence of jobs with explicit gender requests in developing countries, Kuhn et al. (2020) and Chaturvedi et al. (2021) used text analysis and machine learning techniques to predict the implicit maleness and femaleness for individual words in job ads, which provides gendered words about both worker's personalities and required skills. More specifically, Kuhn et al. (2020) apply the naïve Bayesian classifier to identify the likelihood of an explicit gender request based on the words in job titles in a Chinese job board, and Chaturvedi et al. (2021) make use of the text contained in detailed job descriptions in India and construct measures on whether the job ad text is predictive of an employer's explicit male or female preference using a multinomial logistic regression classifier. In this paper, I combine the gendered words from the above studies to create a list of male and female words in my job ads.

As an alternative approach, I carried out two surveys to collect people's perceptions about stereotypically male and female words in job ads. In the English version of the survey, I recruited participants from Amazon Mechanical Turk (MTurk), and let them rate words on maleness and femaleness. A corresponding Chinese version was conducted on Chinese workers. The question takes the same form for each word: "Suppose you are a recruiter and craft a job advertisement containing the following word, you tend to hire (a) no gender requirement, (b) men, (c) women". In this setting, people would perceive the ideal gender of the candidate for jobs that are masculinely or femininely worded. Details on the surveys are provided in Appendix D2 and D3.

The heatmap in Table 7 demonstrates the results on the consistency of stereotypical gender roles in words from these three approaches: previous literature, Mturk survey, and Chinese survey. The displayed male and female words are the ones that were identified by my audit study (from Table 6), and the color intensity represents the femaleness or maleness defined by three approaches. If a word is highlighted with bright red, it is defined as a female word in all three approaches. Words in light red are defined as female words in two approaches, and pink color means a female word in one approach. Male words are marked with blue colors, in which bright blue, light blue and pale blue represent maleness from three, two and one approach, respectively.

Overall, red colors in the left panel and blue colors in the right panel clearly demonstrate that male and female words, which emerge with significantly different frequency in male-only and female-only jobs, are correlated with gender stereotypes. Jobs that are only seen by men contain greater words describing male characteristics, such as engineering, leadership, and overtime, while women are more likely to be exposed to the ads including assist, administrative, patient and temperament, which are also conformed to female stereotypes. The wording in the job descriptions may convey information on the job's implicit gender requests through job recommender systems that encode gender with words reflecting workplace stereotypes on men and women.[27]

[27] This is consistent with results from Chaturvedi et al. (2021), who found that words related to hard-skills and flexibility are critical in explaining gender disparities in labor market outcomes.

Since gendered words occur differentially in gender-specific jobs, I further ask that, among these words, which of them can predict whether the job is a male-only job or a female-only job? I use four methods to measure the relation between words used in job descriptions, and whether it is an only-to-male job.

The first and basic method is OLS regression, in which the outcome variable is 1 if it is a job only recommended to male applicants, and 0 if it is a female-only job. The regressors are dummy variables for the presence of 167 words. Column 1 in Table 8 lists the top 10 words in magnitude that are significant at 5% level. and the overall F-test result is F(167, 24921) =2.63 ($p<0.0001$), indicating that the 167 words are jointly significant. As the matrix is large, sparse, and some of the words are correlated with each other, one may want to select variables that have a larger impact on the outcome rather than including all of them. I use lasso and ridge regressions that impose a penalty parameter for adding an extra variable to figure out which words correspond to the different recommendations to men and women. I applied 20-fold cross-validation to find the optimal penalty parameters, and the selected top 10 words by lasso and ridge regressions are shown in column 2 and 3.

The last approach to identify words that contribute to the classification of jobs recommended to men and women is a machine learning method, random forest. Given the binary measures for the outcome and independent variables, my data structure is very suitable for adopting a decision tree method to find the important factors that affect the sample splitting to male-only and female-only jobs. Column 4 in Table 8 presents the top 10 words with high feature importance based on 100 decision trees and Gini impurity. I find that the three regressions' results are quite consistent. For instance, marriage leave, base pay, words about working hours and breaks are highly predictive of gender-specific job recommendations. Random forest results suggest that words related to breaks, holiday and vacation, are important in making a job ad more or less male recommended.

Finally, to achieve an overall evaluation of gender bias in words, I compute the vector dissimilarity between the average jobs recommended to men and women. Based on the extracted words in the job descriptions, job i can be expressed by a vector S_i with 167 elements, in which the ith word s_{ij} $(j = 1, ..., 167)$ equals 1 if job i contains word j. The dissimilarity between the average male-only job, $\bar{S}^M$ and the average female-only job $\bar{S}^F$ is computed as Euclidean distance between two vectors, and is plotted in Figure 7. It suggests that on the aggregate level, wording in jobs to men and women has a dissimilarity about 0.3, and gender-specific jobs recommended to young workers in male-dominated jobs and entry-level jobs have a slightly higher dissimilarity than such seen by older workers in gender-neutral jobs and middle-level jobs.

7 Explanations for the Gender Bias in Job Recommendations

Section 4 pointed out four mechanisms that could deliver gender-biased job recommendations. In this section, I attempt to distinguish between these reasons, in order to isolate which ones account for that bias.

First, my findings suggest that item-based CF enlarges gender bias in the application process, at least in part. Because the job recommender systems absorb workers' rating behaviors and suggest jobs that are similar to the previously applied jobs, recommended jobs would be more diverse when workers have different application histories. I isolate the impact of item-based CF on gender difference by comparing how the recommended jobs change before and after making applications. Quantitatively, according to Figure 3, the set difference rate rises remarkably after applicants send out profiles in Round 1. Figure 8(a) illustrates the gender gaps of explicit measures on jobs' quality increase after applications, which is particularly true for requested working experience. For the wording in job ads, Figure 8(b) shows that on the aggregate level, word dissimilarity between male- and female-only jobs increases after workers start to apply. Furthermore, by checking the number of male and female words (defined by proportion test with 5% significance level), I find that gendered words in gender-specific jobs increase after applications: The number of male words increases largely from 8 to 19, and the number of female words increases from 18 to 23.

Secondly, the association between gender and words established by content-based recommendations plays a role in gender-biased job recommendations. On the basis of findings from Table 7, female words and male words contained in gender-specific jobs are correlated with gender stereotypes in the workplace. For instance, *figure*, *patient*, and *maternity leave* are feminine-themed words, which also have a higher frequency in the female-only jobs, while jobs recommended to males involve more maleness words such as *engineering* and *leadership*, implying that gender-related words may be encoded and applied into the job recommendations.

Thirdly, rule-based approach that complies with employers' stated gender requests probably has a very limited effect on gender-biased recommendations. While I cannot observe the preferred gender from public job ad postings, recent studies show that jobs advertised specifically for men or women have substantially reduced due to the recent policy interventions from the Chinese government (Kuhn and Shen, 2021). In Kuhn and Shen (2013), jobs that specified desired gender accounted for about 10.5% in Zhaopin in 2008, while my internal data from Liepin suggests that the share was lower than 1% in 2018. Moreover, if the gender requests still exist, they are more likely to appear in the fields that are dominated by one gender, thus we expect to find greater gender bias in male- and female-dominated jobs. However, my decomposition on the setbeahvior difference rate as well as the explicit measures for the quality of gender-specific jobs implies that there is no strong evidence that applicants in male- or female-dominated jobs received more gender-specific job recommendations, or the gender disparities in job's quality magnified in those fields.

Finally, other features of my results suggest that recruiter-behavior based algorithms affect job recommendations. Although my fictitious profiles are very brief and rarely get callbacks from employers, the number of profile views are recorded by websites, including the views from hiring agents who process the received applications, as well as the views from other hiring agents who find the worker's resume through search function or worker recommendations in the job board. If a hiring agent shows interest in a certain worker, then the worker may be recommended to apply for jobs posted by that hiring agent, indicating that the human bias might be manifested and reinforced by the algorithm bias.

To prove this claim, I run the regression of the set difference rate on the number of views on the male and female profiles on gender pair level:

$$Y_i = \beta_0 + \beta_1 ViewF_i + \beta_2 ViewD_i + AX + e_i \tag{4}$$

The outcome variable Y_i is the number of different recommended jobs per 100 recommendations in gender pair i (100*set difference rate), and the variables of interest are the number of views on female profile, $ViewF$, and the gap of received views between female and male profiles in the gender pair, $ViewD$. Table 9 reports the regression results. Column 1 only includes the two measures of views on gender pairs. Column 2 and 3 add controls for worker's age and job's gender type. In column 4, I further control for job board fixed effects to absorb various behaviors of hiring agents in different job boards. The estimation results show that views of the female profile are a significant contributor to the quantity of different jobs seen by identical men and women. The effect of the gender gap in views on the share of gender-specific jobs remains insignificant, however.

It is worth noting that although item-based CF, content-based recommendations, and recruiter-behavior based algorithms potentially generate and perpetuate the gender bias in job recommender systems, I cannot rule out the interactions between those channels, for instance, and several mechanisms can lead simultaneously to the biased results.

8 Conclusion and Discussion

Computer scientists have proposed a variety of ways to improve fairness in algorithms. Most of these approaches focus on enhancing the algorithms' design using computations and formulas to minimize the risk of unfair treatment of certain groups of people (Bozdag, 2013). Recently, a strand of economic studies has attempted to eliminate algorithmic bias by introducing economic concepts into the algorithmic predictions. For instance, Kleinberg et al. (2018) point out that blinding algorithms to the candidate's identity is not a panacea for eliminating biases. They also argue that the inclusion of social planner who cares about equity in the prediction model can promote algorithmic fairness. In studies on particular applications, Arnold et al. (2021) propose approaches to measure discrimination in algorithmic predictions in the context of pretrial bail decisions, and Mullainathan and Obermeyer (2021) consider the label choice bias in algorithms. Moreover, Rambachan et al. (2020) focus on the prediction policy problems and address how to establish the optimal algorithmic regulation from the perspective of economics.

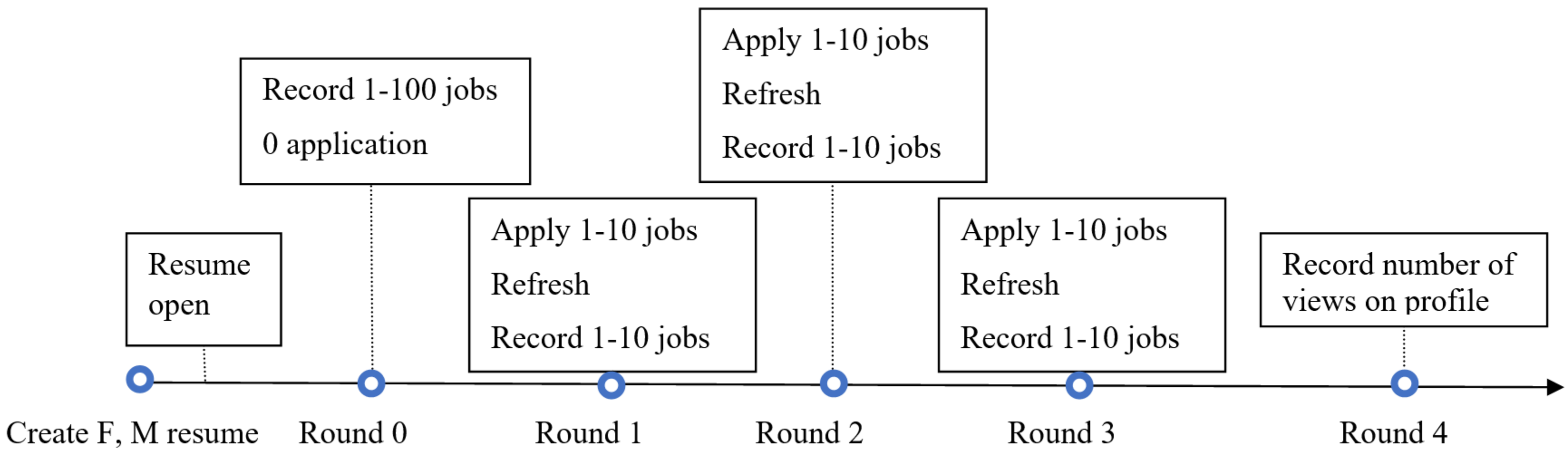
In the specific case of recommender systems, sophisticated recommendation algorithms have proven to be effective in supporting human decisions in discovering new items and are broadly applied in various fields. However, users, sometimes even the designers, have limited knowledge of the recommendation generation process, and the 'black box' of recommender systems may inadvertently cause problems in real social and economic life. However, empirical evidence is still in its infancy in this field. My paper provides an example through an algorithm audit to assess the causal effect of gender in job recommender systems. Using both set differences and list differences to measure the gender gap, I find that identical male and female applicants received different recommendations, in which women were more likely to see low-wage jobs requesting less working experience, requiring literacy and administrative skills, and containing words related to female stereotypes than comparable men. With the growing use of online job searching and recruiting, further research on gender differentials in labor markets should take the job recommendation bias into account.

Since the main objective of the recommendation system is to accurately predict users' (i.e., job seekers on internet job boards) interests, the objective of fairness is possibly overlooked and fails to be incorporated into such systems (Sonboli et al., 2021). While recently some researchers have proposed the fairness-aware recommender systems, it remains an open challenge given that fairness is difficult to define, track, and validate in recommendation systems in which every user expects a different item list based on her taste (Ge et al., 2021; Gao et al., 2021; Beutel et al., 2019; Fu et al., 2020). In my context, job recommender systems that are free of gender bias should theoretically ensure that male and female workers with the same qualifications get recommendations of jobs with the same quality. But what if men and women behave differently in job search such that men are more likely to click optimistically on high-paid jobs than women (Burke et al., 2018)? Furthermore, on *multi-sided* platforms, fairness and utilities of *all* stakeholders should be considered. When hiring agents' feedback influences the recommendation results, should the recommender system truly reflect employers' preference on desirable workers to facilitate potential job-worker matches even if the human decisions are biased? If fairness involves showing workers many jobs they have almost no chance of getting, is that desirable? In the long run, how to maintain the algorithmic fairness when new variables are introduced in the dynamics of job applications? What are the principles to make adjustments or corrections when the bias is detected in the system? All these questions remain unanswered and are good candidates for additional theoretical and empirical research.

Due to data limitations and the high complexity of job recommender systems, it is difficult to find the exact reason or sole driver for the gender bias in job recommendations from the observational data in algorithm audits (Hannák et al., 2017). More importantly, how the gender-biased job recommendations affect job seekers' searching outcomes is still masked. I hope that future research using field experiments or internal data from platforms will shed more light on those questions, and provide additional insights for anti-discrimination policy and legislation.

Tables & Figures

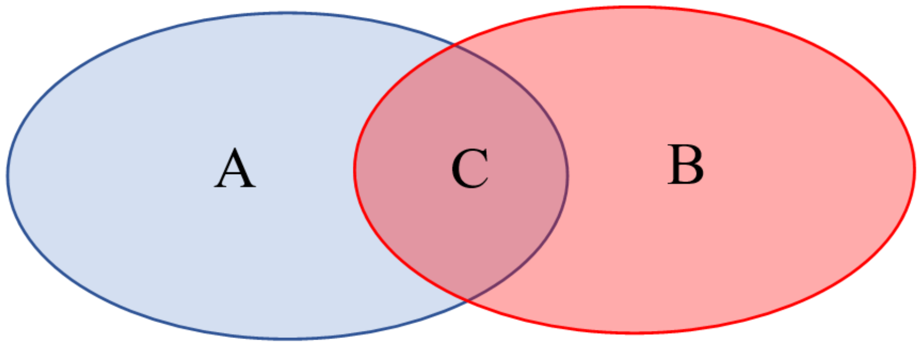
Figure 1: Timeline of the Experimental Steps



Note: Two profiles in each gender pair follow the same timeline. From Round 1 to Round 3, fictitious workers apply for the first job to the 10th job that are displayed in their customized job recommendation interfaces, and the time interval for each round is two weeks.

Figure 2: Difference Measures in Job Recommendations

(a) Set Difference



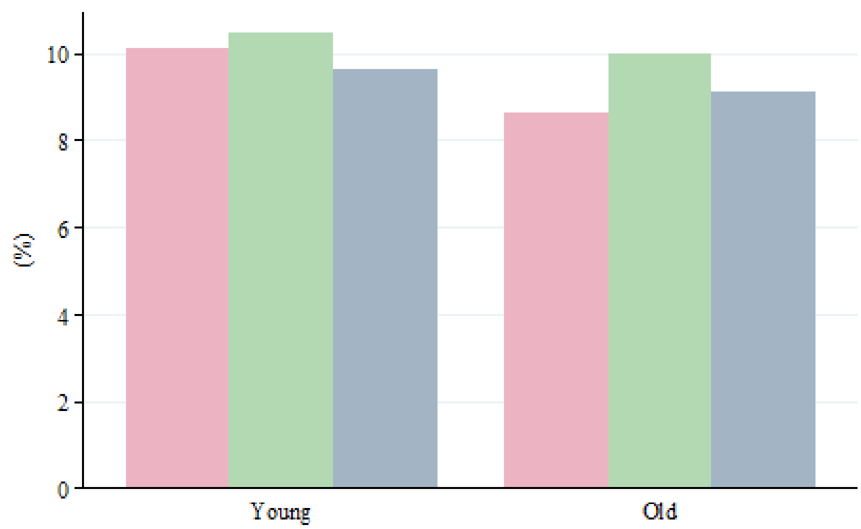
(b) List Difference

	Male	Female	
1st	Job 1	Job 1	Same
2nd	Job 2	Job 2	Same
3rd	Job 3	Job 4	
:			
ith	Job i	Job i+1	
:			
nth	Job n	Job 3	

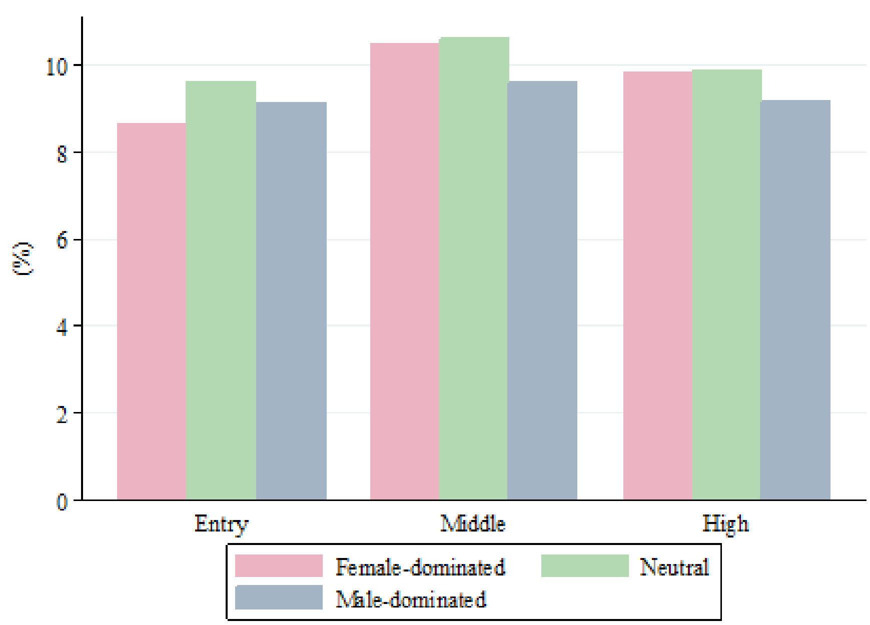
Note: In (a), Set A represents jobs that are only recommended to male applicants, set B represents jobs that are only recommended to female applicants, and set C represents the jobs that are recommended to both males and females. The set difference rate is defined as the share of gender-specific jobs on the complete set of recommended jobs, (A+B)/(A+B+C).
In (b), the shadow area indicates the identical recommendations in the gender pair, in which the ith recommended job in the recommendation list of pairwise male and female applicant is the same. The list difference rate is defined as the share of different recommendations in the recommendation list. In the above example, only the first two jobs in recommendation lists are the same, then list difference rate is (n-2)/n.

Figure 3: Set Difference Rate by Age, Job Gender Type, and Job Hierarchy

(a) Set Difference Rate by Job Gender Type and Age

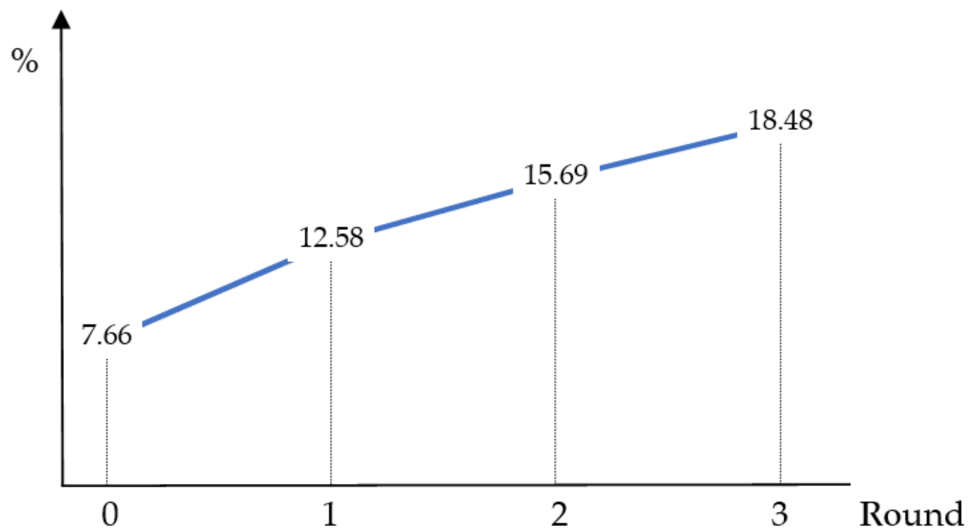


(b) Set Difference Rate by Job Gender Type and Hierarchy



Note: Set difference rate is defined on each group level. For instance, the first bar in (a) is the share of gender-specific jobs on the total jobs recommended to young workers in female-dominated fields.

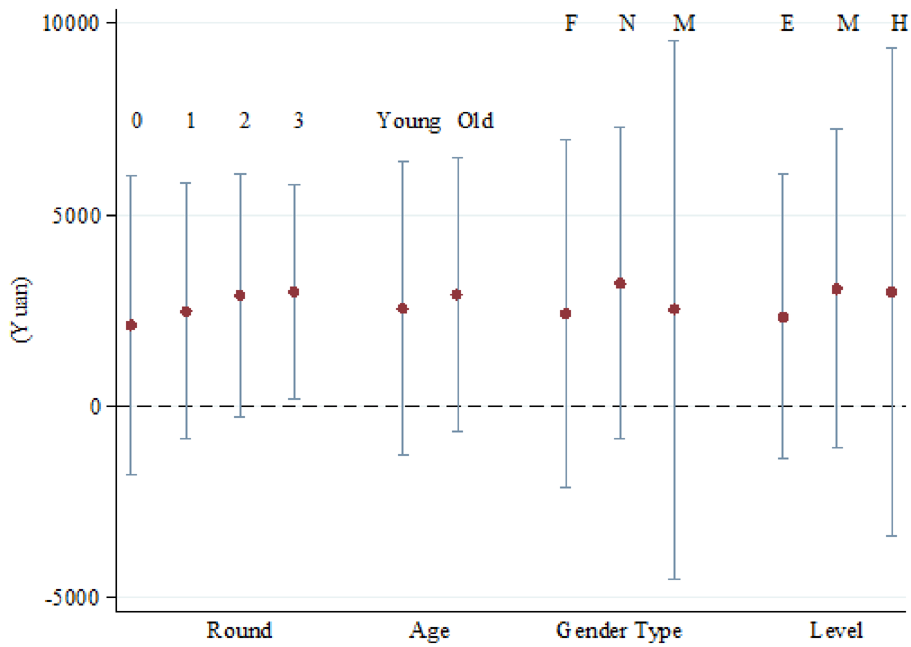
Figure 4: Set Difference Rate by Experimental Rounds



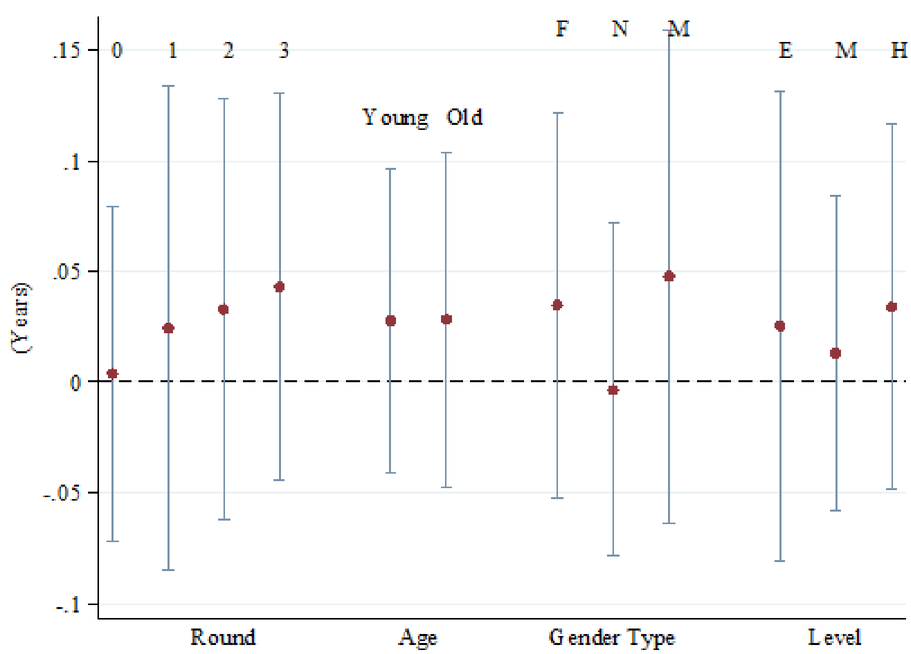
Note: The number of job recommendations in Round 0 to Round 3 is 100, 20, 20, 20, respectively.

Figure 5: Gender Differences on Explicit Measures by Groups

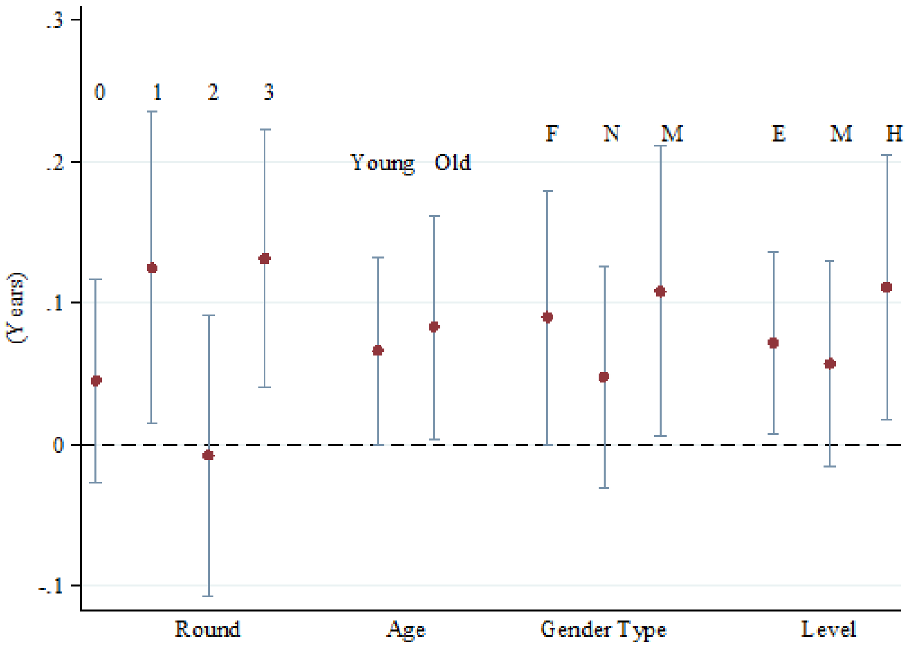
(a) Gender Differences on Posted Wage by Groups



(b) Gender Differences on Requested Education by Groups



(c) Gender Differences on Requested Experience by Groups

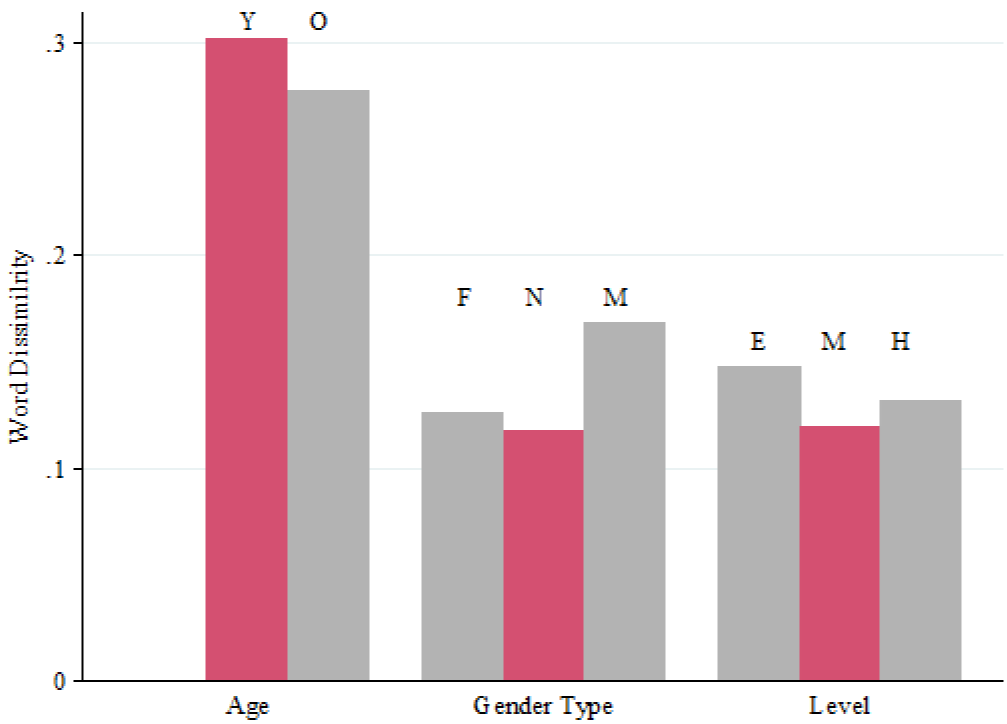


Note: In the job gender type, F denotes female-dominated jobs, N denotes gender-neutral jobs, and M denotes male-dominated jobs. In the job hierarchy level, E denotes entry-level jobs, M denotes middle-level jobs, and H denotes high-level jobs.

Figure 6: Word Cloud from Job Ads

Note: The word cloud is based on the extracted words in the job descriptions from 119,356 recommended job advertisements, and the size corresponds the word frequency. The Chinese version is shown in Appendix Figure D1.

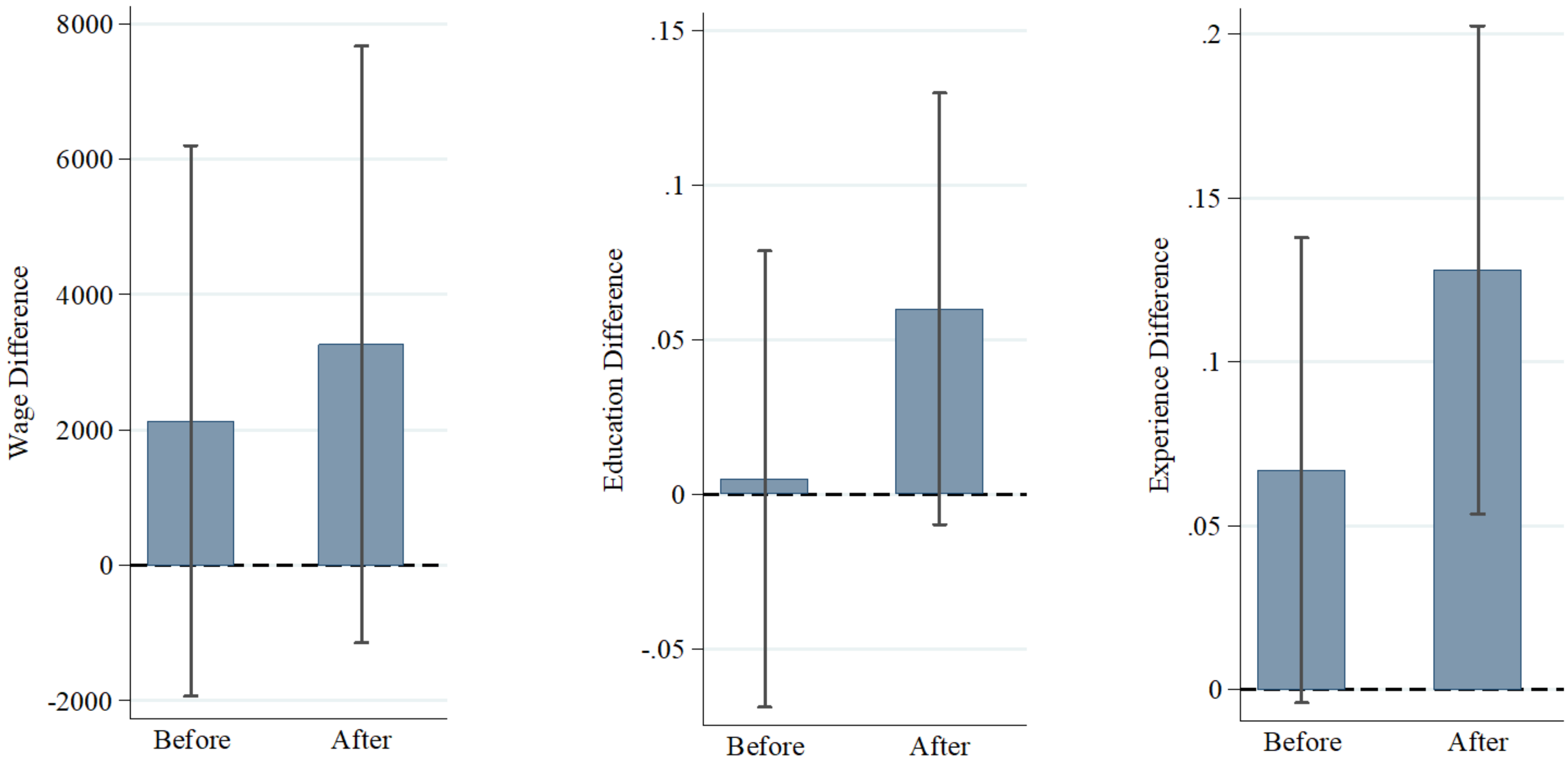
Figure 7: Measure of Words' Dissimilarity in Job Recommendations



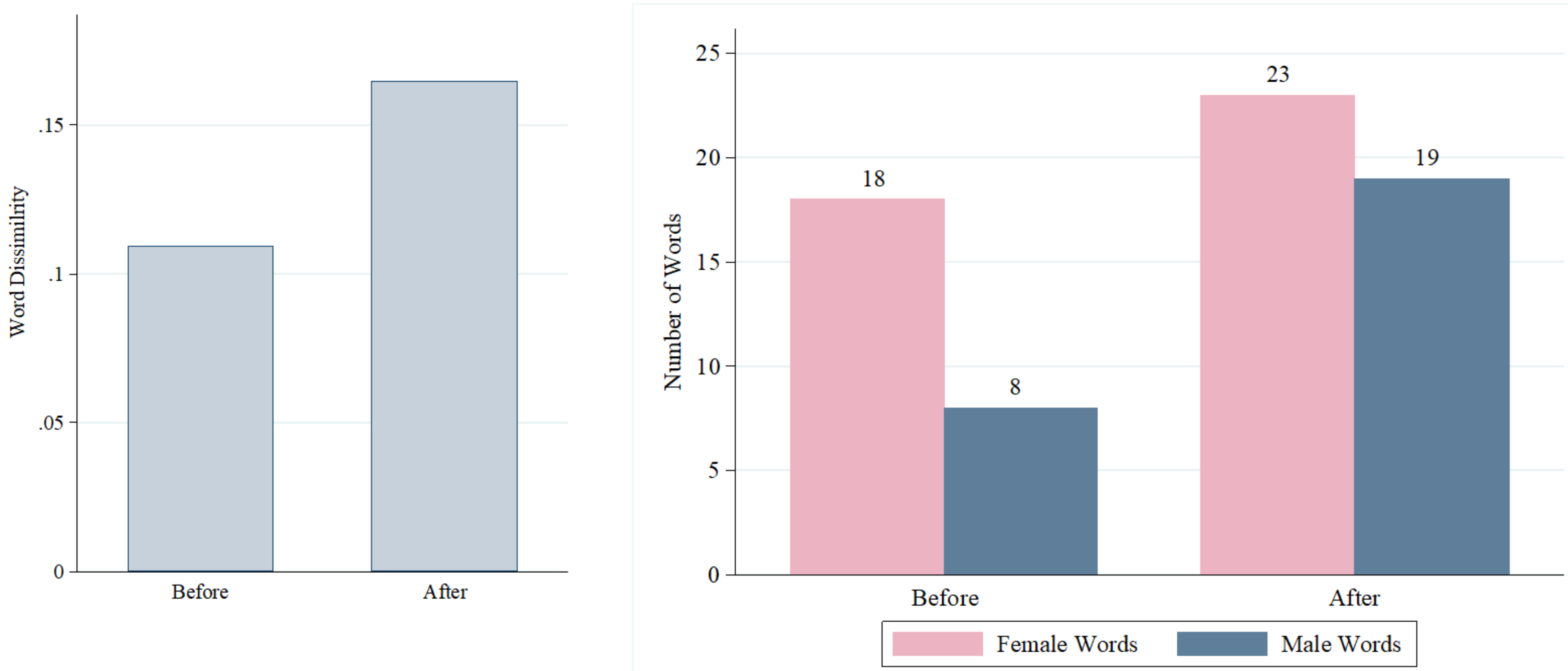
Note: The vector of the average only-to-male (female) jobs consists of 167 elements, in which each element represents the average frequency of that word in the only-to-male (female) sample. Dissimilarity is defined as the Euclidean distance between the male and female vectors.

Figure 8: Comparison of Gender-Specific Jobs Before and After Applications

(a) Comparisons on Explicit Measures



(b) Comparisons on Gender-Specific Words



Note: Figure 8(a) shows the gender gap of explicit measures on gender-specific jobs that are recommended before and after applications separately.
Figure 8(b) compares the wording in gender-specific jobs before and after applications. Word dissimilarity is defined as it in Figure 7, and the number of gendered words in job ads is computed with the same method in Table 6.

Table 1: Descriptive Statistics: Applicant Sample

	Mean
Current Wage	142507.1
	(65141.8)
Desired Wage	179732.1
	(81818.9)
Education	15.56
	(0.4960)
Sample Size	2,240

Note:
1. Current wage and desired wage are annual wage in RMB.
2. Education levels in resumes are transformed to the years of education. A college degree is equivalent to 15 years of education, and a bachelor's degree is equivalent to 16 years of education.
3. Standard errors are in parentheses.

Table 2: Descriptive Statistics: Recommended job Sample

	Mean
Posted Wage?	0.9569
	(0.2032)
Wage, if posted	211004.3
	(658266.8)
Required Education	14.816
	(2.1684)
Required Experience	2.3082
	(2.1119)
Large Company	0.3564
	(0.4789)
Sample Size	119,536

Note:
1. Wage is the midpoint of the posted range of wages.
2. Education levels in job ads are transformed to the years of education. Middle school takes 9 years of education, tech school and high school are equivalent with 12 years of education, college is 15 years of education, and bachelor's degree is equivalent with 16 years of education, master/MBA is 18 years of education, and doctoral degree is 23 years of education.
3. Large company refers to companies that have more than 1,000 employees. The company size is self-reported by hiring agents.
4. Standard errors are in parentheses.

Table 3: Descriptive Statistics: Job Recommendation Sample

	Mean
Desired Wage Match	0.8658
Education Match	0.8812
Experience Match	0.9183
Location Match	0.9924
Sample Size	319,974

Note:
1. Desired wage match equals 1 if the recommended job's upper bound of posted wage range is higher than the worker's lowest desired wage.
2. Education (experience) match is 1 if the worker's years of education (experience) are above the request from the recommended job.
3. Location match is 1 if the worker's city is consistent with the job's city.

Table 4: Set Difference Rate in Job Recommendations

	Share
All	0.0972
Age	
Young	0.1014
Old	0.0929
Gender	
Female	0.0939
Neutral	0.1023
Male	0.0938
Hierarchy	
Entry	0.0902
Middle	0.1029
High	0.0968
City	
Beijing	0.0961
Shanghai	0.1005
Shenzhen	0.0983
Guangzhou	0.0938
Sample Size	25,099

Note:
1. Set difference rate is computed by the number of gender-specific jobs over the number of jobs recommended to both male and female applicants.
2. Duplicates of job recommendations from different rounds are counted once.

Table 5: Gender Differences on Explicit Measures of Recommended Jobs

	Male – Female
Posted Wage	2708.78*
	(1526.03)
Required Education	0.0226
	(0.0259)
Required Experience	0.0779***
	(0.0261)

Note: Gender difference is computed from the mean of male-only jobs minus the mean of female-only jobs. Standard errors are in parentheses, which are derived from two-sample t-tests with equal variance. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Gender Difference in Words in Job Recommendations

	Female Words	Male Words
Skills	assist (-0.0195), data (-0.0195), administrative (-0.0188), speak (-0.0160), chat tools (-0.0122), documentation (-0.0101)	decision making (0.0071), design (0.0102), cooperation (0.0127), teamwork (0.141), engineering (0.0177), manage (0.0177), independent (0.0213)
Work Form	flexible (-0.0176), regular hour (-0.0124), weekly break (-0.0184)	commute (0.0129), work overtime (0.0202), nightwork (0.0084)
Benefits	base pay (-0.0200), marriage leave (-0.0091), maternity leave (-0.0089), social security (-0.0076), unemployment insurance (-0.0021), parental level (-0.0029)	shuttle (0.0139), medical insurance (0.0155), vacation (0.0168), meal (0.0213), reward (0.0256), stock (0.0269)
Company	training (-0.0133)	public company (0.0138)
Requirements	punctual (-0.0128), careful (-0.0125), outgoing (-0.0091), facial (-0.0026), figure (-0.0079), patient (-0.0068), healthy (-0.0064), temperament (0.0063), new grad (-0.0053), below 35 (-0.0050), non experience (-0.0041), trustworthy (-0.0032)	self-motivated (0.0073), pressure (0.0102), innovative (0.0107), experienced (0.0125),

Note: Table 6 displays words that have significantly different probabilities of presenting in male-only and female-only jobs. Coefficients in parentheses represent the gender difference (male-female). Female (male) words are defined from the proportion test with negative (positive) gender differences that are significant at 5% level.

Table 7: Gender Differences on Words and Gender Stereotypes

	Female Words	Male Words
Skills	assist, data, administrative, speak, chat tools, documentation	decision making, design, cooperation, teamwork, engineering, manage, independent, leadership
Work Form	flexible, regular hour, weekly break	commute, work overtime, nightwork
Benefits	base pay, marriage leave, maternity leave, social security, unemployment insurance, parental level	shuttle, medical insurance, vacation, meal, reward, stock
Company	training	public company
Requirements	punctual, careful, outgoing, facial figure, patient, healthy, temperament, new grad, below 40, non experience, trustworthy	self-motivated, pressure, innovative, experienced

Note: Table 7 shows the relation between gendered words in job ads and gender stereotypes. The color intensity indicates the maleness and femaleness consistency with gender stereotypes from literature and two survey results. Female words are highlighted with red colors, male words highlighted with blue colors, and strong color indicates high consistency.

Table 8: Top 10 Words in Prediction of Gender-Specific Recommended Jobs

OLS	Lasso	Ridge	Random Forest
night work	base pay	night work	activities
trustworthy	regular hour	trustworthy	meal
below40	marriage leave	work shift	commute
base pay	engineering	big and small week	holiday
regular hour	independent	facial	vacation
data	stock	marriage leave	commission
flexible	public	below40	allowance
documentation	overtime	endowment ins	training
weekly break	flexible	maternity leave	fiveone
administrative	data	base pay	reward

Note:
1. Table 8 presents the top words in predicting whether a job is only recommended to male applicants. The outcome variable is binary and equals 1 for male-only jobs, and independent variables are 167 dummy variables for the existence of words in job ads.
2. Column 1 lists words from the OLS regression, which are significant at 5% level and sorted in descending order of the magnitude of coefficients.
3. Column 2 and 3 present words that are selected by the Lasso and Ridge regression. The penalty parameter for Lasso regression is 0.23 and is 0.31 in Ridge regression. Those are determined by using 20-fold cross-validation for the highest R squared. Words are sorted in descending order of the magnitude of estimation effects.
4. In column 4, random forest is applied to find words that have high impacts on the classification of male-only and female-only jobs based on 100 bootstraps and Gini impurity. Words are sorted in descending order of the importance factor.
5. fiveone represents "five social insurance and one housing fund" (五险一金), including endowment insurance, medical insurance, unemployment insurance, employment injury insurance, maternity insurance and housing fund. Big and small week describes the working schedule in which workers have one-day rest in one week and two-day rest in the next week.

Table 9: Effects of Views from Hiring Agents on Job Recommendations

	(1)	(2)	(3)	(4)
ViewF	0.0141*	0.0188***	0.0165**	0.0203***
	(0.007)	(0.007)	(0.007)	(0.007)
ViewD	-0.0010	-0.0008	0.0046	0.0072
	(0.013)	(0.013)	(0.013)	(0.013)
Age		Yes	Yes	Yes
Job Gender Type			Yes	Yes
Job Board				Yes
N	1033	1033	1033	1033
R^2	0.0109	0.0212	0.0408	0.0633

Note:
1. The dependent variable is the number of gender-specific jobs in 100 recommended jobs.
2. In column 1, the regressors are the number of views on the female's profile and the gender gap on the number of views (male-female) in each gender pair. Column 2 controls for young or older pairs. Column 3 further controls the worker's job gender type, including female-dominated jobs, gender-neutral jobs and male-dominated jobs. Column 4 adds the job board fixed effect.
3. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

References

Al-Otaibi, S. T. and Ykhlef, M. (2012), 'A survey of job recommender systems', *International Journal of Physical Sciences* **7**(29), 5127–5142.

Albert, R., Escot, L. and Fernández-Cornejo, J. A. (2011), 'A field experiment to study sex and age discrimination in the madrid labour market', *The International Journal of Human Resource Management* **22**(02), 351–375.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A. and Rieke, A. (2019), 'Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes', *Proceedings of the ACM on Human-Computer Interaction* **3**(CSCW), 1–30.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016), 'Machine bias', *ProPublica, May* **23**(2016), 139–159.

Arnold, D., Dobbie, W. and Hull, P. (2021), Measuring racial discrimination in algorithms, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 49–54.

Athreya, B. (2021), 'Bias in, bias out: gender and work in the platform economy'.

Avivi, H., Kline, P., Rose, E. and Walters, C. (2021), Adaptive correspondence experiments, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 43–48.

Baert, S. (2018), 'Hiring discrimination: An overview of (almost) all correspondence experiments since 2005', *Audit studies: Behind the scenes with theory, method, and nuance* pp. 63–77.

Baert, S., Albanese, A., du Gardein, S., Ovaere, J. and Stappers, J. (2017), 'Does work experience mitigate discrimination?', *Economics Letters* **155**, 35–38.

Banfi, S., Choi, S. and Villena-Roldan, B. (2019), 'Deconstructing job search behavior', *Available at SSRN 3323545* .

Barocas, S. and Selbst, A. D. (2016), 'Big data's disparate impact', *Calif. L. Rev.* **104**, 671.

Bartlett, R., Morse, A., Stanton, R. and Wallace, N. (2021), 'Consumer-lending discrimination in the fintech era', *Journal of Financial Economics* .

Barzilay, A. R. and Ben-David, A. (2016), 'Platform inequality: gender in the gig-economy', *Seton Hall L. Rev.* **47**, 393.

Bem, S. L. (1981), 'Gender schema theory: A cognitive account of sex typing.', *Psychological review* **88**(4), 354.

Bertrand, M. and Duflo, E. (2017), 'Field experiments on discrimination', *Handbook of economic field experiments* **1**, 309–393.

Bertrand, M., Goldin, C. and Katz, L. F. (2010), 'Dynamics of the gender gap for young professionals in the financial and corporate sectors', *American economic journal: applied economics* **2**(3), 228–55.

Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E. H. et al. (2019), Fairness in recommendation ranking through pairwise comparisons, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', pp. 2212–2220.

Black, S. E. and Spitz-Oener, A. (2010), 'Explaining women's success: technological change and the skill content of women's work', *The Review of Economics and Statistics* **92**(1), 187–194.

Booth, A. and Leigh, A. (2010), 'Do employers discriminate by gender? a field experiment in female-dominated occupations', *Economics Letters* **107**(2), 236–238.

Bozdag, E. (2013), 'Bias in algorithmic filtering and personalization', *Ethics and information technology* **15**(3), 209–227.

Brenčič, V. (2012), 'Wage posting: evidence from job ads', *Canadian Journal of Economics/Revue canadienne d'économique* **45**(4), 1529–1559.

Burke, R., Sonboli, N. and Ordonez-Gauger, A. (2018), Balanced neighborhoods for multi-sided fairness in recommendation, *in* 'Conference on Fairness, Accountability and Transparency', PMLR, pp. 202–214.

Bustelo, M., Díaz Escobar, A. M., Lafortune, J., Piras, C., Salas, L. M. and Tessada, J. A. (2020), What is the price of freedom? estimating women's willingness to pay for job schedule flexibility, Technical report, IDB Working Paper Series.

Caliskan, A., Bryson, J. J. and Narayanan, A. (2017), 'Semantics derived automatically from language corpora contain human-like biases', *Science* **356**(6334), 183–186.

Capéau, B., Eeman, L., Groenez, S., Lamberts, M. et al. (2012), 'Two concepts of discrimination: inequality of opportunity versus unequal treatment of equals', *Ecore Discussion Papers* **58**, 127.

Card, D., Cardoso, A. R. and Kline, P. (2016), 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women', *The Quarterly Journal of Economics* **131**(2), 633–686.

Carlsson, M. (2011), 'Does hiring discrimination cause gender segregation in the swedish labor market?', *Feminist Economics* **17**(3), 71–102.

Cediey, E. and Foroni, F. (2008), *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office*, ILO.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J. and Mullainathan, S. (2016), 'Productivity and selection of human capital with machine learning', *American Economic Review* **106**(5), 124–27.

Chaturvedi, S., Mahajan, K., Siddique, Z. et al. (2021), Words matter: Gender, jobs and applicant behavior, Technical report, IZA DP No. 14497.

Chen, L., Ma, R., Hannák, A. and Wilson, C. (2018), Investigating the impact of gender on rank in resume search engines, *in* 'Proceedings of the 2018 chi conference on human factors in computing systems', pp. 1–14.

Christl, M. and Köppl-Turyna, M. (2020), 'Gender wage gap and the role of skills and tasks: evidence from the austrian piaac data set', *Applied Economics* **52**(2), 113–134.

Cook, C., Diamond, R., Hall, J. V., List, J. A. and Oyer, P. (2021), 'The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers', *The Review of Economic Studies* **88**(5), 2210–2238.

Correll, S. J., Benard, S. and Paik, I. (2007), 'Getting a job: Is there a motherhood penalty?', *American journal of sociology* **112**(5), 1297–1338.

Cortés, P., Pan, J., Pilossoph, L. and Zafar, B. (2021), Gender differences in job search and the earnings gap: Evidence from business majors, Technical report, National Bureau of Economic Research.

Cowgill, B. (2018), 'Bias and productivity in humans and algorithms: Theory and evidence from resume screening', *Columbia Business School, Columbia University* **29**.

Craswell, N., Zoeter, O., Taylor, M. and Ramsey, B. (2008), An experimental comparison of click position-bias models, *in* 'Proceedings of the 2008 international conference on web search and data mining', pp. 87–94.

de Sá, C. C., Gonçalves, M. A., Sousa, D. X. and Salles, T. (2016), Generalized broof-l2r: a general framework for learning to rank based on boosting and random forests, *in* 'Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval', pp. 95–104.

Dreisbach, C., Koleck, T. A., Bourne, P. E. and Bakken, S. (2019), 'A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data', *International journal of medical informatics* **125**, 37–46.

Eriksson, S. and Lagerström, J. (2012), 'The labor market consequences of gender differences in job search', *Journal of Labor Research* **33**(3), 303–327.

Faberman, R. J. and Kudlyak, M. (2019), 'The intensity of job search and search duration', *American Economic Journal: Macroeconomics* **11**(3), 327–57.

Fitzpatrick, M. A., Mulac, A. and Dindia, K. (1995), 'Gender-preferential language use in spouse and stranger interaction', *Journal of Language and Social Psychology* **14**(1-2), 18–39.

Fix, M., Galster, G. C. and Struyk, R. J. (1993), 'An overview of auditing for discrimination', *Clear and convincing evidence: Measurement of discrimination in America* **1**, 13–14.

Flory, J. A., Leibbrandt, A. and List, J. A. (2015), 'Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions', *The Review of Economic Studies* **82**(1), 122–155.

Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., Xu, S., Geng, S., Shah, C., Zhang, Y. et al. (2020), Fairness-aware explainable recommendation over knowledge graphs, *in* 'Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval', pp. 69–78.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A. (2020), 'Predictably unequal? the effects of machine learning on credit markets', *The Effects of Machine Learning on Credit Markets (October 1, 2020)* .

Gaddis, S. M. (2018), An introduction to audit studies in the social sciences, *in* 'Audit studies: Behind the scenes with theory, method, and nuance', Springer, pp. 3–44.

Gao, R., Ge, Y. and Shah, C. (2021), 'Fair: Fairness-aware information retrieval evaluation', *arXiv preprint arXiv:2106.08527* .

Gastil, J. (1990), 'Generic pronouns and sexist language: The oxymoronic character of masculine generics', *Sex roles* **23**(11), 629–643.

Gaucher, D., Friesen, J. and 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.

Ge, Y., Liu, S., Gao, R., Xian, Y., Li, Y., Zhao, X., Pei, C., Sun, F., Ge, J., Ou, W. et al. (2021), Towards long-term fairness in recommendation, *in* 'Proceedings of the 14th ACM International Conference on Web Search and Data Mining', pp. 445–453.

Gee, L. K. (2019), 'The more you know: Information effects on job application rates in a large field experiment', *Management Science* **65**(5), 2077–2094.

Hannák, A., Wagner, C., Garcia, D., Mislove, A., Strohmaier, M. and Wilson, C. (2017), Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr, *in* 'Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing', pp. 1914–1933.

He, H., Neumark, D. and Weng, Q. (2021), 'Do workers value flexible jobs? a field experiment', *Journal of Labor Economics* **39**(3), 709–738.

Helleseter, M. D., Kuhn, P. and Shen, K. (2020), 'The age twist in employers' gender requests evidence from four job boards', *Journal of Human Resources* **55**(2), 428–469.

Hoffman, C. and Hurst, N. (1990), 'Gender stereotypes: Perception or rationalization?', *Journal of personality and social psychology* **58**(2), 197.

Hong, W., Zheng, S., Wang, H. and Shi, J. (2013), 'A job recommender system based on user clustering.', *J. Comput.* **8**(8), 1960–1967.

Jannach, D., Resnick, P., Tuzhilin, A. and Zanker, M. (2016), 'Recommender systems—beyond matrix completion', *Communications of the ACM* **59**(11), 94–102.

Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010), *Recommender systems: an introduction*, Cambridge University Press.

Jiang, R., Chiappa, S., Lattimore, T., György, A. and Kohli, P. (2019), Degenerate feedback loops in recommender systems, *in* 'Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society', pp. 383–390.

Jiang, S., Hu, Y., Kang, C., Daly Jr, T., Yin, D., Chang, Y. and Zhai, C. (2016), Learning query and document relevance from a web-scale click graph, *in* 'Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval', pp. 185–194.

Kay, M., Matuszek, C. and Munson, S. A. (2015), Unequal representation and gender stereotypes in image search results for occupations, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', pp. 3819–3828.

Keith, K. and McWilliams, A. (1999), 'The returns to mobility and job search by gender', *ILR Review* **52**(3), 460–477.

Kellogg, K. C., Valentine, M. A. and Christin, A. (2020), 'Algorithms at work: The new contested terrain of control', *Academy of Management Annals* **14**(1), 366–410.

Kim, P. T. (2017), 'Auditing algorithms for discrimination', *U. Pa. L. Rev. Online* **166**, 189.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V. and Jain, A. K. (2012), 'Face recognition performance: Role of demographic information', *IEEE Transactions on Information Forensics and Security* **7**(6), 1789–1801.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Rambachan, A. (2018), Algorithmic fairness, *in* 'Aea papers and proceedings', Vol. 108, pp. 22–27.

Köchling, A. and Wehner, M. C. (2020), 'Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of hr recruitment and hr development', *Business Research* pp. 1–54.

Kuhn, P. J. and Shen, K. (2021), What happens when employers can no longer discriminate in job ads?, Technical report, National Bureau of Economic Research.

Kuhn, P. and Shen, K. (2013), 'Gender discrimination in job ads: Evidence from china', *The Quarterly Journal of Economics* **128**(1), 287–336.

Kuhn, P., Shen, K. and Zhang, S. (2020), 'Gender-targeted job ads in the recruitment process: Facts from a chinese job board', *Journal of Development Economics* **147**, 102531.

Lambrecht, A. and Tucker, C. (2019), 'Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads', *Management science* **65**(7), 2966–2981.

Lee, D. H. and Brusilovsky, P. (2007), Fighting information overflow with personalized comprehensive information access: A proactive job recommender, *in* 'Third International Conference on Autonomic and Autonomous Systems (ICAS'07)', IEEE, pp. 21–21.

Lee, M. K. and Baykal, S. (2017), Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division, *in* 'Proceedings of the 2017 acm conference on computer supported cooperative work and social computing', pp. 1035–1048.

Li, D., Raymond, L. R. and Bergman, P. (2020), Hiring as exploration, Technical report, National Bureau of Economic Research.

Lindqvist, A., Renström, E. A. and Sendén, M. G. (2019), 'Reducing a male bias in language? establishing the efficiency of three different gender-fair language strategies', *Sex Roles* **81**(1), 109–117.

Mann, G. and O'Neil, C. (2016), 'Hiring algorithms are not neutral', *Harvard Business Review* **9**, 2016.

Marinescu, I. and Wolthoff, R. (2020), 'Opening the black box of the matching function: The power of words', *Journal of Labor Economics* **38**(2), 535–568.

Mas, A. and Pallais, A. (2017), 'Valuing alternative work arrangements', *American Economic Review* **107**(12), 3722–59.

Mullainathan, S. and Obermeyer, Z. (2021), On the inequity of predicting a while hoping for b, *in* 'AEA Papers and Proceedings', Vol. 111, pp. 37–42.

Nosek, B. A., Banaji, M. R. and Greenwald, A. G. (2002), 'Harvesting implicit group attitudes and beliefs from a demonstration web site.', *Group Dynamics: Theory, Research, and Practice* **6**(1), 101.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science* **366**(6464), 447–453.

OECD (2016), *The survey of adult skills: Reader's companion*, Organisation for Economic Co-operation and Development.
URL: *https://doi.org/10.1787/9789264204027-en*

Parsons, D. O. (1991), 'The job search behavior of employed youth', *The Review of Economics and Statistics* pp. 597–604.

Pedreshi, D., Ruggieri, S. and Turini, F. (2008), Discrimination-aware data mining, *in* 'Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 560–568.

Pekkarinen, T. and Vartiainen, J. (2006), 'Gender differences in promotion on a job ladder: Evidence from finnish metalworkers', *ILR Review* **59**(2), 285–301.

Petit, P. (2007), 'The effects of age and family constraints on gender hiring discrimination: A field experiment in the french financial sector', *Labour Economics* **14**(3), 371–391.

Pető, R. and Reizer, B. (2021), 'Gender differences in the skill content of jobs', *Journal of Population Economics* **34**(3), 825–864.

Poutanen, S. and Kovalainen, A. (2017), New economy, platform economy and gender, *in* 'Gender and Innovation in the New Economy', Springer, pp. 47–96.

Prassl, J. (2018), *Humans as a service: The promise and perils of work in the gig economy*, Oxford University Press.

Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2020), Mitigating bias in algorithmic hiring: Evaluating claims and practices, *in* 'Proceedings of the 2020 conference on fairness, accountability, and transparency', pp. 469–481.

Rambachan, A., Kleinberg, J., Ludwig, J. and Mullainathan, S. (2020), An economic perspective on algorithmic fairness, *in* 'AEA Papers and Proceedings', Vol. 110, pp. 91–95.

Rambachan, A. and Roth, J. (2019), 'Bias in, bias out? evaluating the folk wisdom', *arXiv preprint arXiv:1909.08518* .

Riach, P. A. and Rich, J. (2006), 'An experimental investigation of sexual discrimination in hiring in the english labor market', *Advances in Economic Analysis & Policy* **5**(2).

Rich, J. (2014), 'What do field experiments of discrimination in markets tell us? a meta analysis of studies conducted since 2000'.

Richardson, M., Dominowska, E. and Ragno, R. (2007), Predicting clicks: estimating the click-through rate for new ads, *in* 'Proceedings of the 16th international conference on World Wide Web', pp. 521–530.

Roberts, D. C. and Utych, S. M. (2020), 'Linking gender, language, and partisanship: Developing a database of masculine and feminine words', *Political Research Quarterly* **73**(1), 40–50.

Rudman, L. A. and Kilianski, S. E. (2000), 'Implicit and explicit attitudes toward female authority', *Personality and social psychology bulletin* **26**(11), 1315–1328.

Sandvig, C., Hamilton, K., Karahalios, K. and Langbort, C. (2014), 'An algorithm audit', *Data and Discrimination: Collected Essays. Washington, DC: New America Foundation* pp. 6–10.

Savage, D. D. and Bales, R. (2016), 'Video games in job interviews: using algorithms to minimize discrimination and unconscious bias', *ABAJ Lab. & Emp. L.* **32**, 211.

Siting, Z., Wenxing, H., Ning, Z. and Fan, Y. (2012), Job recommender systems: a survey, *in* '2012 7th International Conference on Computer Science & Education (ICCSE)', IEEE, pp. 920–924.

Sonboli, N., Smith, J. J., Cabral Berenfus, F., Burke, R. and Fiesler, C. (2021), Fairness and transparency in recommendation: The users' perspective, *in* 'Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization', pp. 274–279.

Stinebrickner, T. R., Stinebrickner, R. and Sullivan, P. J. (2018), Job tasks and the gender wage gap among college graduates, Technical report, National Bureau of Economic Research.

Tambe, P., Cappelli, P. and Yakubovich, V. (2019), 'Artificial intelligence in human resources management: Challenges and a path forward', *California Management Review* **61**(4), 15–42.

Yu, H., Liu, C. and Zhang, F. (2011), 'Reciprocal recommendation algorithm for the field of recruitment', *Journal of Information & Computational Science* **8**(16), 4061–4068.

Zhou, X., Zhang, J. and Song, X. (2013), 'Gender discrimination in hiring: Evidence from 19,130 resumes in china', *Available at SSRN 2195840* .

Appendices

Appendix A

Resume Audit Study Experimental Design

A1: Job Type Selection

In each job board, 35 types of jobs were selected based on three criteria: the number of active job openings, the job's gender type (female-dominated jobs, gender-balanced jobs, and male-dominated jobs), and hierarchy level (entry, middle, and high). For each job type, I scraped 50 job ads to determine the education level and academic major that are required by most employers. In addition, 50 resumes in the job type were employed to derive the current wages (adjusted to be age-appropriate).

Table A1 lists the selected job type (industry-occupation cell) in each job board, the corresponding hierarchy level (low, middle, high), the required education level, and the major. Current wage (,) represents current wages for (young, older) workers in 10k RMB, respectively.

Table A1.1 Selected Job Types in Job Board 1

Gender	Industry	Occupation	Hierarchy Level	Education Level	Major	Current Wages
M	Computer Software	Software Engineer	Low	Bachelor	Computer Science	(14, 17)
	Computer Software	Senior Software Engineer	High	Bachelor	Computer Science	(17, 23)
	Internet/ E-Business	Operations Specialist	Low	College	Computer Science	(7, 9)
	Internet/ E-Business	Operations Manager/Supervisor	High	Bachelor	Computer Science	(11, 14)
	Machine Manufacturing	General Worker /Operator		College	Machinery	(8, 13)
	Automobiles/Motorcycles	General Worker /Operator		College	Machinery	(9, 13)
	Transportation/Shipping	Courier		College	Econ&Management	(5, 6)
	Internet/ E-Business	Courier		College	Econ&Management	(6, 7)
	Wholesale/Retail	Warehouse Keeper		College	Econ&Management	(4, 5)
N	Internet/ E-Business	Data Analyst		Bachelor	Statistics	(11, 14)
	Computer Software	Data Analyst		Bachelor	Statistics	(11, 14)
	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(13, 17)
	Internet/ E-Business	Product Manager/Supervisor		Bachelor	Econ&Management	(13, 17)
	Internet/ E-Business	Sales Representative	Low	College	Marketing	(5, 7)
	Education/Training	Sales Representative	Low	College	Marketing	(5, 7)
	Real Estate Services	Sales Representative	Low	College	Marketing	(6, 8)
	Internet/ E-Business	Sales Manager	Middle	College	Marketing	(12, 17)
	Computer Software	Sales Manager	Middle	College	Marketing	(12, 17)
	Wholesale/Retail	Sales Director	High	Bachelor	Marketing	(16, 21)
	Internet/ E-Business	Sales Director	High	Bachelor	Marketing	(16, 21)
	Internet/ E-Business	Front Desk	Low	College	Econ&Management	(6, 8)

F	Professional Services	Front Desk	Low	College	Econ&Management	(6, 8)
	Professional Services	Executive Assistant	Low	College	Econ&Management	(7, 9)
	Computer Software	Executive Assistant	Low	College	Econ&Management	(7, 9)
	Internet/ E-Business	Executive Manager	High	College	Econ&Management	(11, 13)
	Wholesale/Retail	Store Clerk	Low	College	Marketing	(5, 7)
	Wholesale/Retail	Store Manager	High	College	Marketing	(9, 11)
	Internet/ E-Business	Customer Service	Low	College	Marketing	(5, 6)
	Finance/Securities	Customer Service	Low	College	Marketing	(5, 6)
	Internet/ E-Business	Customer Service Manager	High	College	Marketing	(8, 12)
	Trade/Import-Export	Accountant		Bachelor	Accounting	(8, 12)
	Wholesale/Retail	Accountant		Bachelor	Accounting	(8, 12)
	Internet/ E-Business	HR Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Professional Services	HR Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Internet/ E-Business	Human Resources Manager	High	College	Econ&Management	(9, 12)

Table A1.2 Selected Job Types in Job Board 2

Gender	Industry	Occupation	Skill Level	Education Level	Major	Current Wages
M	Computer Software	Software Engineer		Bachelor	Computer Science	(15, 23)
	Internet	Mobile Development Engineer		Bachelor	Computer Science	(16, 23)
	Internet	Algorithm Engineer		Bachelor	Computer Science	(17, 24)
	Internet	Operations Specialist	Low	College	Computer Science	(7, 9)
	Internet	Operations Manager/Supervisor	High	Bachelor	Computer Science	(11, 14)
	Real Estate Development	Real Estate Project Management		Bachelor	Architecture	(14, 22)
N	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(14, 20)
	Internet	Product Manager/Supervisor		Bachelor	Econ&Management	(14, 20)
	Computer Software	Project Manager/Supervisor		Bachelor	Econ&Management	(13, 19)
	Internet	Project Manager/Supervisor		Bachelor	Econ&Management	(13, 19)
	Internet	Data Analyst		Bachelor	Statistics	(12, 18)
	Big Data	Data Analyst		Bachelor	Statistics	(12, 18)
	Securities/Investment	Data Analyst		Bachelor	Statistics	(12, 18)
	Advertising/Public Relations	Public Relations Specialist/Assistant		College	Marketing	(11, 14)
	Advertising/Public Relations	Public Relations Manager/Supervisor		Bachelor	Marketing	(15, 20)
	E-Business	Sales Representative	Low	College	Marketing	(7, 12)
	Internet	Sales Representative	Low	College	Marketing	(7, 12)
	Education/Training	Sales Representative	Low	College	Marketing	(7, 12)
	Real Estate Services	Sales Representative	Low	College	Marketing	(8, 13)
	Wholesale/Retail	Sales Manager	Middle	College	Marketing	(12, 17)
	Real Estate Services	Sales Manager	Middle	College	Marketing	(12, 17)

	Internet	Sales Director	High	Bachelor	Marketing	(14, 19)
	Wholesale/Retail	Sales Director	High	Bachelor	Marketing	(14, 19)
F	E-Business	Web Customer Service	Low	College	Marketing	(6, 8)
	Banking	Telephone Customer Service	Low	College	Marketing	(6, 8)
	E-Business	Customer Service Manager	High	College	Marketing	(12, 15)
	Banking	Customer Service Manager	High	College	Marketing	(12, 15)
	E-Business	Accountant		Bachelor	Accounting	(9, 14)
	Internet	HR Specialist/Assistant	Low	College	Econ&Management	(6, 9)
	Professional Services	HR Specialist/Assistant	Low	College	Econ&Management	(6, 9)
	Internet	Human Resources Manager/Supervisor	High	Bachelor	Econ&Management	(11, 14)
	Computer Software	Human Resources Manager/Supervisor	High	Bachelor	Econ&Management	(11, 14)
	Internet	Executive Assistant/Secretary	Low	College	Econ&Management	(7, 9)
	Internet	Administration Specialist/Assistant	Low	College	Econ&Management	(6, 8)
	Internet	Administration Manager/Supervisor	High	College	Econ&Management	(9, 14)

Table A1.3 Selected Job Types in Job Board 3

Gender	Industry	Occupation	Skill Level	Education Level	Major	Current Wages
M	Internet/E-Business	WEB Front-end Developer		Bachelor	Computer Science	(17, 24)
	Machine Manufacturing	Mechanical Engineer		Bachelor	Machinery	(16, 21)
	Computer Software	Software Engineer	Low	Bachelor	Computer Science	(18, 25)
	Computer Software	Senior Software Engineer	High	Bachelor	Computer Science	(22, 27)
	Internet/E-Business	Operations Specialist	Low	College	Computer Science	(10, 13)
	Internet/E-Business	Operations Manager/Supervisor	High	Bachelor	Computer Science	(14, 20)
	Real Estate Development	Architect		Bachelor	Architecture	(15, 22)
N	Pharmaceuticals/Biotechnology	Sales Representative	Low	College	Marketing	(10, 15)
	Securities/Investment Funds	Sales Representative	Low	College	Marketing	(11, 15)
	Pharmaceuticals/Biotechnology	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(14, 18)
	Internet/E-Business	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 18)
	Securities/Investment Funds	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 18)
	Pharmaceuticals/Biotechnology	Sales Director	High	Bachelor	Marketing	(17, 24)
	Internet/E-Business	Sales Director	High	Bachelor	Marketing	(16, 25)
	Commodity	Sales Director	High	Bachelor	Marketing	(16, 24)
	Internet/E-Business	Product Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Computer Software	Product Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Internet/E-Business	Project Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Computer Software	Project Manager/Supervisor		Bachelor	Econ&Management	(15, 22)
	Commodity	Marketing Manager/Supervisor		Bachelor	Marketing	(14, 22)
	Wholesale/Retail	Marketing Manager/Supervisor		Bachelor	Marketing	(14, 22)

	Real Estate Development	Legal manager/Supervisor		Bachelor	Law	(15, 25)
	Internet/E-Business	Legal manager/Supervisor		Bachelor	Law	(15, 24)
F	Internet/E-Business	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Real Estate Development	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Internet/E-Business	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Real Estate Development	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Internet/E-Business	Human Resources Director	High	Bachelor	Econ&Management	(16, 26)
	Real Estate Development	Human Resources Director	High	Bachelor	Econ&Management	(16, 26)
	Internet/E-Business	Accountant	Low	Bachelor	Accounting	(12, 18)
	Securities/Investment Funds	Financial Manager	High	Bachelor	Finance	(15, 20)
	Internet/E-Business	Administration Specialist/Assistant	Low	College	Econ&Management	(9, 13)
	Real Estate Development	Executive Assistant/Secretary	Low	College	Econ&Management	(10, 14)
	Internet/E-Business	Administration Manager/Supervisor	Low	Bachelor	Econ&Management	(15, 20)
	Internet/E-Business	Administration Vice President	High	Bachelor	Econ&Management	(51, 88)

Table A1.4 Selected Job Types in Job Board 4

Gender	Occupation	Skill Level	Education Level	Major	Current Wages
M	WEB Front-end Developer		Bachelor	Computer Science	(19, 25)
	Operation and Maintenance Engineer	Low	Bachelor	Computer Science	(18, 24)
	Operation and Maintenance Director	High	Bachelor	Computer Science	(19, 26)
	Pattern Recognition		Bachelor	Computer Science	(19, 25)
	Machine Learning		Bachelor	Computer Science	(19, 25)
	Operations Assistant	Low	College	Computer Science	(7, 9)
	Operations Specialist	Middle	College	Computer Science	(10, 12)
	Operations Manager/Supervisor	High	Bachelor	Computer Science	(14, 19)
	Test Engineer	Low	Bachelor	Computer Science	(15, 22)
	Test Manager	High	Bachelor	Computer Science	(19, 25)
	Data Architect		Bachelor	Computer Science	(17, 25)
N	Sales Representative	Low	College	Marketing	(8, 12)
	Sales Manager/Supervisor	Middle	Bachelor	Marketing	(13, 17)
	Sales Director	High	Bachelor	Marketing	(18, 25)
	Product Assistant	Low	College	Econ&Management	(9, 10)
	Product Manager	High	Bachelor	Econ&Management	(15, 23)
	Project Assistant	Low	College	Econ&Management	(9, 10)
	Project Manager	High	Bachelor	Econ&Management	(15, 23)
	Data Analyst		Bachelor	Statistics	(13, 19)
	Design Assistant	Low	College	Arts	(8, 10)
	Designer	Middle	College	Arts	(13, 19)

	Design Manager	High	Bachelor	Arts	(15, 23)
	Strategy Consultant		Bachelor	Econ&Management	(13, 19)
F	Human Resources Specialist/Assistant	Low	College	Econ&Management	(9, 10)
	Human Resources Manager/Supervisor	Middle	Bachelor	Econ&Management	(14, 20)
	Human Resources Director	High	Bachelor	Econ&Management	(17, 26)
	Accountant	Low	Bachelor	Accounting	(13, 17)
	Training Specialist		College	Econ&Management	(10, 12)
	Customer Service	Low	College	Marketing	(7, 8)
	Customer Service Manager	High	College	Marketing	(13, 17)
	Media Specialist	Low	College	Marketing	(7, 8)
	Media Manager	High	Bachelor	Marketing	(10, 15)
	Administration Specialist/Assistant	Low	College	Econ&Management	(9, 12)
	Administration Manager/Supervisor	Middle	Bachelor	Econ&Management	(13, 18)
	Administration Director	High	Bachelor	Econ&Management	(16, 25)

Note: The industry in job board 4 is set as “all industries”.

A2. Fictitious Resume

The resumes only contain the basic information required by each job board to register as a valid job seeker. The first section of a fictitious resume is personal information, including worker's name, birth date, years of working experience, current wage, city, employment status, phone number, and email address. The second part is about worker's education: the highest education level, time period, university name and major. The third part describes worker's working experience of the most recent job including the time period, company name, occupation, industry, job title, and job description. The last part is worker's intention for future jobs, including desired wage, desired location, desired industry, and occupation. Two workers in each gender pair have identical backgrounds, and four workers in each group (young male, young female, older male, older female) are placed in each job type.

A2.1 Personal Information

Name: I picked up the most popular first and last names to make up the names of fictitious applicants. Based on the statistics from 2015 Chinese Census 1% Population Sample, I chose the top 20 last names, top 15 male first names, and top 15 female first names as the applicants' name pool (listed in Appendix A2.1). For each applicant, the last name and first name corresponding to the applicant's gender will be randomly drawn from the name pool. Although gender is explicitly stated in the resume and we do not need applicant's name to denote gender, I still adopted first names that are consistent with a worker's gender to make the fictitious profile as common and real as possible.

Names of Fictitious Applicants

Last name: 李(Li), 王(Wang), 张(Zhang), 刘(Liu), 陈(Chen), 杨(Yang), 赵(Zhao), 黄(Hunag), 周(zhou), 吴 (Wu), 徐(Xu), 孙(Sun), 胡(Hu), 朱(Zhu), 高(Gao), 林(Lin), 何(He), 郭(Guo),马(Ma), 罗(Luo).

Male First Name: 伟(Wei), 强(Qiang), 磊(Lei), 军(Jun), 洋(Yang), 勇(Yong), 杰(Jie), 涛(Tao), 超(Chao), 平(Ping), 刚(Gang), 浩(Hao), 鹏(Peng), 宇(Yu), 明(Ming).

Female First Name: 芳(Fang), 娜(Na), 敏(Min), 静(Jing), 丽(Li), 艳(Yan), 娟(Juan), 霞(Xia), 婷(Ting), 雪(Xue), 丹(Dan), 英(Ying), 洁(Jie), 玲(Ling), 燕(Yan).

Birth Date: Employers infer worker's age from the birth date. Instead of varying workers' age directly, I used their graduation year to classify the age level, and "older" workers refer to ones who graduated earlier and have more working experience. Applicants have two potential age levels: Young workers graduated in 2017, and old workers graduated in 2007. After a worker's graduation year is fixed, his age is jointly determined by the graduation year and his education level. The advantage of this design is that workers' years of working experience are equalized within each age level. More specifically, young workers are 25 (with a college degree, born in 1995) or 26 (with a bachelor's degree, born in 1994) with three years of working experience, 35 or 36 years old are for the senior workers with more than 5 years of working experience. Workers in the gender pair have the same randomly drawn birth month and day.

Years of Working Experience: To simplify the profiles, I assumed workers started to work just after they graduated from the university/college of their highest degree. As discussed above, years of working experience is the difference between the current year (2020) and the graduation year. For instance, if a worker graduated in 2017, then he has 2020 – 2017, three years of working experience.

Current Wage: Fictitious workers' wages are drafted based on wages of active workers in job boards by matching their current job position as well as working experience. I used the hiring agent account in each platform and searched for workers that were currently in the job positions and specified the working experience as "1 to 3 years" and "5 to 10 years" in March 2020. For each experience level in every job position, I recorded the first 50 workers' current wages shown in the search result and took the average as the fictitious worker's wage.

City: All of the four job boards are nationally recognized and cover most of the regions in China, and over half of job postings are from first-tier cities. To achieve enough amount of job recommendations, fictitious workers are currently living in the first-tier cities, including Beijing, Shanghai, Shenzhen, and Guangzhou.

Employment Status: All of the workers are currently employed.

Phone number and email: Each applicant has a unique and active email address and mobile phone number.

A2.2 Education

Workers' education level is designed to match jobs' education requirements. For each job type, I checked 100 job advertisements in February 2020 and listed the most common education. 85% of job ads required workers had a bachelor's or junior college degree. Bachelor's degree often takes 4 years to achieve, while junior college takes 3 years. The end time of school is the graduation year, and the start time of school depends on worker's education degree, which is three years (college degree) or four years (bachelor's degree) earlier than the graduation year. For instance, a young worker, graduated with a bachelor's degree in June 2017, is 26 years old (born in 1994) and started his university program in August 2013.

Two workers in the same gender pair have the same educational background, and the school's name is randomly drawn from the Chinese High Education Institution List, released by the Ministry of Education in 2019. Majors will also match job positions: Computer Science/Software is for IT jobs, Mathematics/Statistics is for data position, and economics/management/marketing majors are for other jobs.

A2.3 Recent Job History

As we assume all the workers are currently employed, their recent jobs are their current jobs. For young workers, their current jobs started in August in the year when they graduated with the highest degree (2017); for old workers, their current jobs started five years ago, in March 2015, implying that they have 5 years tenure in their recent positions.

I made up company names to minimize the disturbance to both job seekers and employers on job platforms. The company name consists of three parts: (1) company's location. It will be the same with worker's current city. (2) company's name. I used an online business name generator to collect 100 company names listed below. The company name will be randomly assigned to each gender pair. (3) company's industry. It will be consistent with the job's industry. An example of the company name is, Beijing Dongya Internet Technology Company.

Worker's current occupation and industry will be the same as the job's occupation and industry. Job title and job description are filled in by words, and I set them as the job's occupation.

Names of Company

东艾, 森利, 先卓, 利晟, 同通, 富长盛, 芯达, 精典, 尼佳, 益复捷, 生德, 晶长, 森益, 金伙伴, 德光, 茂全, 鲜派, 信顺康, 龙丝, 新耀协, 佳丽, 昇晖, 佳洲, 森道尔, 皇祥千, 润飞昌, 福中荣, 基玉, 如和, 茂乾, 翔鹏, 南湘, 圣泰, 吉春, 本寿, 亚义金, 耀浩, 邦洁, 宝复, 洪进贵, 永泰满, 显郦, 华行, 韵仪, 格派, 晶佩, 迪和, 领速, 贝耀, 信华诚, 世力, 舜杰, 久福, 曼新, 仁大兴, 金祥元, 泰伟飞, 亚和金, 吉振, 和伟中, 盛金缘, 立韦, 宏久, 吉至, 曼展, 天联, 金涛, 网诚, 系广, 圣金龙, 易露发, 嘉利华, 聚顿, 公同宏, 威邦, 力涛, 恒蓝, 铭航, 中美公, 永逸, 同捷, 发和, 易龙, 汉金, 干亚, 翔洋, 新都, 茂进永, 达通, 娇罗, 浩中和, 东升, 龙姿, 隆新弘, 仟顺, 越福, 川实, 中协吉, 霸辉, 洪谦, 裕飞

A2.4. Job Intention

A worker searches for full-time jobs, in which the desired wage is 120% of his current wage (or the wage range), and the desired city, industry, and occupation will be the same as the current ones.

Table A2.1 summarizes the information included in worker's resume.

Table A2.1: Resume Information Generation

	Method	Note
Personal Information		
Name	Randomly assigned to each worker	Appendix A2.1
Birth Date	Young worker graduated in 2017, and older worker graduated in 2007. Birth year is decided by graduation year and education level.	Young, bachelor's =1994, Young, college=1995. Older, bachelor's =1984, Older, college=1985.
Years of Working Experience	2020 - graduation year	3 or 13 years
Current Wage	Average wage of the collected workers in the platforms.	Adjust with job type and experience.
City	Beijing, Shanghai, Shenzhen, Guangzhou	
Employment Status	Currently employed.	
Phone Number & Email	Uniquely assigned for each worker.	
Education		
Highest degree	Assigned on group level, based on job type's education requirement.	Bachelor's degree or junior college.
Time Period	Graduation year – years to achieve the highest degree.	4 years to achieve bachelor's degree, 3 years to achieve college degree.
School Name	Randomly drawn for each gender pair.	Chinese High Education Institution List (2019)
Major	Same on group level.	Depends on job type.
Recent Job		
Time Period	Young worker: after graduation (2017) until now,	

	Older worker: 2015 until now.	
Company Name	Location +name + industry, name will be randomly assigned to each worker.	Appendix A2.2
Occupation	Same with job type	
Industry	Same with job type	
Job Title	Same with occupation	
Job Description	Same with occupation	
Intention		
Desired Wage	Current wage*1.2	
Desired City	Same with city	
Desired Industry	Same with job type	
Desired Occupation	Same with job type	

Appendix B

Robustness Check on Set Difference of Job Recommendations

Table B1: Gender Differences on Explicit Measures of Recommended Jobs

	Male – Female	
	Two-sample t-test with Unequal Variance	Wilcoxon Rank-sum Test
Posted Wage	2708.78*	2.0474**
	(1527.46)	
Required Education	0.0226	0.8415
	(0.0259)	
Required Experience	0.0779***	3.5723***
	(0.0262)	

Note:

1. Gender difference is computed from the mean of male-only jobs minus the mean of female-only jobs.
2. In column 2, standard errors are in parentheses, which are derived from two-sample t-tests with unequal variance.
3. Column 3 reports z-value from Wilcoxon Rank-sum Test.
4. *** p<0.01, ** p<0.05, * p<0.1.

Appendix C

List Difference of Job Recommendations

Table C1: List Difference Rate in Job Recommendations

	Share
All	0.7069
Round	
0	0.5827
1	0.8642
2	0.8634
3	0.8643
Age	
Young	0.7048
Old	0.7092
Gender	
Female	0.7039
Neutral	0.7093
Male	0.7079
Hierarchy	
Entry	0.7049
Middle	0.7077
High	0.7083
City	
Beijing	0.7018
Shanghai	0.7090
Shenzhen	0.7104
Guangzhou	0.7067

Note: List difference is defined as the share of different recommendations in the recommendation list, as shown in Figure 2(b).

Table C2: Gender Differences on Explicit Measures of Recommended Jobs

	Male – Female	
	Paired t-test	Wilcoxon Signed Rank Test
Posted Wage	1207.55	1.3348
	(830.11)	
Required Education	0.0084	0.7739
	(0.0093)	
Required Experience	0.0042	0.5489
	(0.0090)	

Note:

1. Gender difference is computed from the mean of male-only jobs minus the mean of female-only jobs.
2. In column 2, standard errors are in parentheses, which are derived from paired-sample t-tests.
3. Column 3 reports z-value from Wilcoxon Signed Rank Test.
4. *** p<0.01, ** p<0.05, * p<0.1.

Appendix D

Words in Job Recommendations

Figure D1: Word Cloud in Chinese

Table D1: Word List in Job Ads

Skills	Literacy skills: listen, speak, read, write, language, documentation Numeracy skills: data, accounting, analysis ICT skills: programing, microsoft office, chat tools Problem solving skills: learning, comprehension, thinking, logic, decision-making, planning, problem-solving, engineering, independent Influencing skills: leadership, team management, charge, supervise Co-operative skills: cooperation, communication, teamwork, assist, coordination, organize, negotiate, public relation, marketing, sale, client, compliance Self-organizing skills: administrative, design, collect, reception, driving, execution, test, task management
Work Form	Schedule: work shift, night work, morning work, evening work, big and small week*, eight-hour, flexible, attendance, overtime, no overtime Business travel: regular travel, short travel, long travel Work break: weekly break, monthly break, noon break, regular working hour
Benefits	Payment: base pay, commission, stock, allowance, promotion, reward Break: vacation, marriage leave, parental leave, maternity leave, sick leave, funeral leave, holiday Facilities: office supplements, vehicle, meal, housing, shuttle, nearby, commute Insurance: fiveone*, medical insurance, commercial insurance, social security, funds, maternity insurance, unemployment insurance, endowment insurance, injury insurance, disease insurance Other benefits: training, staffing, activities, mentor
Company	Environment: atmosphere, employee care, career, dream, culture, screening Type: direct recruiting, public company, top500, startup, flat, financing, big company* Title: senior, medium, core

Requirements	Education: non education, certificate, new grad, tongzhao*, tier-one school, fulltime school, top school, nonmajor, major, science&engineering Experience: non experience, experienced, oversea Demographics: non gender, non age, below35, below40 Personality: effective, rigorous, carefully, patient, energetic, active, outgoing, optimistic, virtuous, trustworthy, honest, practical, self-motivated, hardworking, passion, tenacious, sharp mind, generous, curious, courageous, innovative, punctual, entrepreneurial, devotion, enthusiasm, kind, responsibility, pressure Appearance: figure, temperament, healthy, facial, clothing, shape Objective: voice, responsive, no crime, regulation, solitary

Note:

1. Table D1 shows the extracted words from job ads in four job boards, and the restrictions are described in Section 6.2.2.
2. Every listed word includes its variations on parts of speech, such as leadership vs leading, and confidence vs confident.
3. fiveone represents “five social insurance and one housing fund” (五险一金), including endowment insurance, medical insurance, unemployment insurance, employment injury insurance, maternity insurance, and housing fund. Big and small week describes the working schedule in which workers have one-day rest in one week and two-day rest in the next week. Big company indicates companies that have more than 1000 employees. Tongzhao means university or college admission is through Gaokao in high school.

D2: Survey from Amazon MTurk

To determine the gendered perceptions of words, I recruited participants from Amazon's Mechanical Turk (MTurk) in September 2021 to choose whether the existence of a certain word in the job ad indicates gender stereotypes and implicit gender preferences of employers.

The survey question is: "Suppose you are the hiring agent of a company, and plan to post a job advertisement that contains the word X in the job description. This indicates that you prefer to hire (1) no gender request for worker; (2) male worker; (3) female worker".

In total, 86 valid surveys were collected from people between the ages of 25 to 55, and 56% of them were men. The gender score of a word is computed as:

Score = -1*number of participants choose (3) + 1* number of participants choose (2)

, in which -1 indicates the extreme female word and 1 implies the extreme male word. The average gender score of words in the survey is 0.0905 and the standard deviation is 0.1111. Male words are defined as words whose scores are above one standard deviation from the mean, 0.2016, and female words' scores are below one standard deviation from the mean, -0.0206.[1]

Table D2: Gendered Words from Amazon MTurk Survey

Female Words	Male Words
administrative, assist, careful, compliance, design, documentation, enthusiasm, figure, holiday, kind, learning, marriage leave, maternity insurance, maternity leave, parental leave, patient, read, reception, shape, sick leave, temperament, voice, writing	analysis, big week, commission, data, driving, responsibility, effective, engineering, evening work, experienced, independent, no crime, leadership, logic, long travel, mentor, negotiation, nightwork, overtime, practical, pressure, promotion, science&engineering, startup, stock, supervise, vehicle, work shift

[1] tierone university and tongzhao are excluded from the surveyed words because they are only identified in the Chinese high-level education system.

D3: Survey from Chinese Workers

The Chinese version of the survey on people's perceptions about gendered words in job ads was conducted in Wenjuanxing (问卷星) in September 2021. The surveyed question is the same as the one from AMturk, but in Chinese: 假设您是公司 HR，发布的招聘广告中包含以下词汇，代表您倾向于招聘 （1 ）性别不限；（2）男员工；（3）女员工。

79 valid respondents participated in the survey, 81% of them were between 25 to 55 years old and 73% of them were men. The average gender score of words in the survey is 0.0962 and the standard deviation is 0.0721. Male words are defined as words whose scores are above one standard deviation from the mean, 0.1683, and female words' scores are below one standard deviation from the mean, 0.0241.

Table D3: Gendered Words from Chinese Survey

Female Words	Male Words
active, administrative, assist, atmosphere, care, collect, communication, compliance, design, eight hour, facial, figure, flexible, health, kind, listen, marriage leave, maternity ins, office supplements, outgoing, parental leave, passion, patient, read, reception, shape, sick leave, speak, temperament, voice, writing	below40, charge, commission, commute, core, courageous, culture, data, disease ins, driving, responsibility, engineering, enterprise, independent, injury ins, no crime, leadership, long travel, meal, negotiation, nightwork, optimistic, oversea, overtime, practical, pressure, promotion, responsive, screening, self-motivated, solving, staffing, stock, teamwork, tenacious, training, unemployment ins, punctual