Bias in Black Saber Hiring and Internal Practices

Examining whether gender has an undue influence on AI scoring, advancement, and salary

Alan Yue

2021-04-21

Contents

Executive Summary

Background & Aims

We were provided with data on both applicants who underwent the hiring process and on all current employees at Black Saber. Our goal was to examine this data to see if there is any latent and unfair bias in either the AI system that oversees the hiring process or the internal processes that determine salary and advancement. We only examine if there was gender bias, as this is the only information that was a potential source of discrimination that we were granted access to. Every person had listed as their gender either 'Man', 'Woman', or 'Prefer not to say'.

Figure 1: Proportion of applicants who made the next round by gender and previous work experience

Hiring

- We examined each phase of the hiring process individually. The first two phases were screened solely by the AI system. The third stage consists of two interviewers scoring the candidate. After that, offers are made.
- Figure 1 is a visualization of some of the data we had to work with. It shows the proportion of applicants who made it from Round 1 to Round 2 by gender and previous work experience

scores. We see that work experience does appear to be a factor in determining whether an applicant moves onto the next round, but gender doesn't appear to be.

- For Phase 1, the AI has never moved a candidate to Phase 2 if they didn't submit either a cover letter or CV. The candidates offered a job were those that achieved the top interview scores in Phase 3. Thus we conclude that these processes are not influenced by gender.
- We modeled the effect of gender on the odds of moving from Phase 1 to Phase 2, independent of the effects of GPA, extracurricular score, work experience score, and the team that was applied for. We found no evidence of gender having an effect on these odds.
- We modeled the effect of gender on the odds of moving from Phase 1 to Phase 2, independent of their technical skills score, writing skills score, speaking skills score, leadership presence score, and the team they applied for. All scores were given by the AI system. We found no evidence of gender having an effect on these odds.
- We modeled the effect of gender on mean interview score from Phase 3, independent of all other factors previously listed. We found no evidence of gender having an effect on these scores.

Promotions

- We modeled the effect of gender on the number of promotions received, independent of the effects of the number of financial quarters worked, team, role seniority when the employee started, and mean productivity and leadership scores over all quarters worked. We found no evidence of gender having an effect on the number of promotions received.

Salary

- We modeled the effect of gender on salary, independent of the effects of productivity score, leadership level, role seniority, and team. We are 95% confident that the mean annual salary of female employees is between \$1700.99 and \$2776.37 less than that of their male counterparts, independent of the other predictors.

Conclusion

We found no evidence of an undue influence of gender on the AI system that conducts the initial hiring process, the interview scores in the third stage of hiring, or the number of promotions given to an employee. We did find evidence to support that the women at Black Saber are paid on average less than men. We must conclude by emphasizing that stating we did not find any evidence of gender discrimination in a particular process does not imply we found evidence indicating there was none.

Technical Report

Introduction

We were provided with data on both applicants who underwent the hiring process and on all current employees at Black Saber. Our goal was to examine this data to see if there is any latent and unfair bias in either the AI system that oversees the hiring process or the internal processes that determine salary and advancement. This report will cover our efforts to answer the research questions given below, including the data wrangling we did, how we chose our models, the results of our analysis, and the conclusions we draw.

Research questions

1. Is the hiring process unfairly influenced by gender such that whether an applicant moves on in each phase of the hiring process is in part solely due to the gender of the applicant?
2. Does gender have an unfair impact on the promotion decisions made such that the number of promotions an employee receives is in part solely due to the gender of the employee?
3. Does gender have an unfair impact on the salary decisions made such that the salary of an employee in a given financial quarter is in part solely due to the gender of the employee?

We deem an influence to be unfair when it is based on factors unrelated to perceived talent or value to the company. Thus for each question, we will conclude an unfair influence is present if it persists in the face of controlling for other factors that are based on talent and value. In other words, if we find it is likely that Black Saber's processes treat different gender groups differently regardless of confounding factors, we will say that process is unfair.

Is the hiring process unfairly influenced by the gender of the applicant?

Data Exploration & Wrangling

To answer this research question, we used the hiring data we were given access to. This included information on each phase of the hiring process and which applicants moved on in each phase. For the first two phases, we had information on the scores given to the applicants by the AI system on all the criteria they were assessed. The third phase information consisted of the scores given by the two interviewers of each candidate in that phase. We also had access to the list of applicants who were offered a job in the end and the gender of each applicant, if it was provided.

We first checked to see that there were no missing values in any of the data we were sent. Then we converted the gender, cover letter, CV, and team applied for variables to factors, as these variables all correspond to different categories the applicants belong in, rather than a numerical score for example. For the cover letter and CV, each applicant is given either a 0 or 1 depending on whether or not one was received; this is not a score on the quality of what was received.

For Phase 3 of the hiring process, we created a new variable: the mean of the two individual interview scores and used this in our analysis. The reason for this is because we were not given access to who the individual interviewers were and were told that the order of the scores (first or second for each applicant) was irrelevant, so we have no way of meaningfully comparing the two. Finally for each phase, we created an indicator variable that would indicate whether the applicant moved on to the next phase of the hiring process or was offered a job in the case of Phase 3.

Figure 1: Proportion of applicants who made the next round by gender and previous work experience

Figure 1 is a visualization of some of the data we had to work with. It shows the proportion of applicants who made it from Round 1 to Round 2 by gender and previous work experience scores. We see that work experience does appear to be a factor in determining whether an applicant moves onto the next round, but gender doesn't appear to be. We will explore this more in depth in the next section.

Methods

We decided to examine each phase of the hiring process individually, to see if any discrimination occurred at any level. We assume our dataset satisfies the property that one applicant being in our dataset is not dependent on a different applicant being in our dataset, so we assume independence between our observational units.

For the first phase, we wanted to model the influence of gender on whether an applicant moved on in each phase, controlling for other relevant variables. Because whether an applicant moved on is a binary outcome, we decided to use binary logistic regression for the first two phases. We initially attempted to use gender, GPA, extracurricular score, work experience score, the team they applied for, whether or not they submitted a cover letter, and whether or not they submitted

a CV as predictors. However this led to the model being able to perfectly predict whether some applicants moved on. After some exploration, we discovered that whenever either the cover letter or CV variables indicated one had not been submitted, the applicant did not move on to Phase 2. Thus we removed these variables from our model because they are redundant; we know exactly what their impact is on whether an applicant moves to Phase 2.

For the second phase, we modelled whether an applicant moved on to Phase 3 with binary logistic regression based on their gender, technical skills score, writing skills score, speaking skills score, leadership presence score, and the team they applied for. We again include this last predictor because we don't know how the AI system treats applicants differently based on the team they applied for. The dispersion parameters for both this and the previous model is taken to be 1, suggesting no evidence of overdispersion. Because under a true model, the residual deviance should follow a chi-squared distribution, we can run a goodness of fit test for both models. We achieved P-values of 0.5543507 and ~1 for the first and second models, respectively. These P-values represent the probability of observing our residual deviances on our residual degrees of freedom if we had true models. Thus we saw no evidence suggesting a lack of fit.

For the third phase, we initially thought to model whether an applicant was finally offered a job with binary logistic regression based on their gender and mean interview score but after further exploration, we saw that the 10 applicants offered a job were simply the 10 that had the highest mean interview score. So instead we decided to model the mean interview score for each applicant based on all previous predictor variables in an effort to discern whether the interviewers themselves were discriminating based on gender. This essentially answers the same question we sought to answer because we assume final job offer decisions are made solely based on the top interview scores. Because this score is continuous, we used a standard linear model with a Gaussian response. We looked at the diagnostic plots for this model and found no serious violations of the linear response, normality or nonconstant variance assumptions.

Results

The coefficients for the first model are given below in Table 1. Because logistic regression uses a logit link function, we can interpret the coefficient on the gender explanatory variables by exponentiating: eˆ-0.0622=0.94 and eˆ-0.6=0.549, indicating that although there's not much evidence that women were treated differently at this stage, the odds of moving onto Phase 2 for those who prefered not to state their gender was slightly more than half that of men, controlling for their GPA, extracurricular score, work experience score, and team they applied for. However we see that the P-values for both these coefficients are far from statistically significant, 0.76 and 0.45 respectively.

Table 1: Phase 1-2 Model Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-6.4411427	0.5445420	-11.8285517	0.0000000
Gender: Women	-0.0622265	0.2036815	-0.3055087	0.7599787
Gender: Prefer not to say	-0.6000902	0.7977601	-0.7522189	0.4519194
GPA	1.9682446	0.2115345	9.3046028	0.0000000
Extracurriculars	0.2733225	0.1971131	1.3866280	0.1655552
Work Experience	0.7615525	0.2570846	2.9622638	0.0030539
Applied for Data Team	0.1013233	0.2064103	0.4908829	0.6235092

We can exponentiate the same way with the second model and get eˆ-0.784=0.456, indicating that the odds of women moving onto the next phase were now less than half that of men, controlling for their technical skills score, writing skills score, speaking skills score, leadership presence score, and team they applied for. We do not obtain a statistically significant result (P-value=0.304), with the confidence interval (0.093293, 1.953969) after exponentiating containing both values below and above 1. This suggests that it's plausible the odds of moving onto the next phase for women could be both less than or more than that of men. We do not interpret the coefficient for those who preferred not to state their gender for this model because there are so few of those applicants in this dataset; the coefficient would be unreliable.

For the interview model, we obtain a coefficient of 0.434 for women. Because this model has a Gaussian response, this should indicate that women receive a higher mean interview score, controlling for all other predictors explored thus far. However our P-value is again too high (0.81) for this to be definitive. There were no applicants who preferred not to state their gender and made it to Phase 3.

Does the gender of the employee have an impact on whether or not they receive promotions?

Data Exploration & Wrangling

To answer this research question, we used the employee data we were given access to. This included the gender of every current employee as well as their team, role seniority, productivity score, a description of their demonstrated leadership, and salary for each financial quarter in which they worked.

We first checked to see that there were no missing values. We converted the salary variable so that it could be interpreted numerically and then checked a number of assumptions were true. We iterated through the employee data and checked that every employee's team never changed, that the quarters they worked had no breaks (they worked every quarter from when they were employed to the present day), and that their seniority levels either remained the same or increased by one level from one quarter to the next (there were never any demotions or promotions by more than one seniority level). These assumptions allow us to disregard the effect that switching teams may have on promotions, disregard the effect of non-continuous work on promotions, and more easily calculate the number of promotions an employee had during their employment.

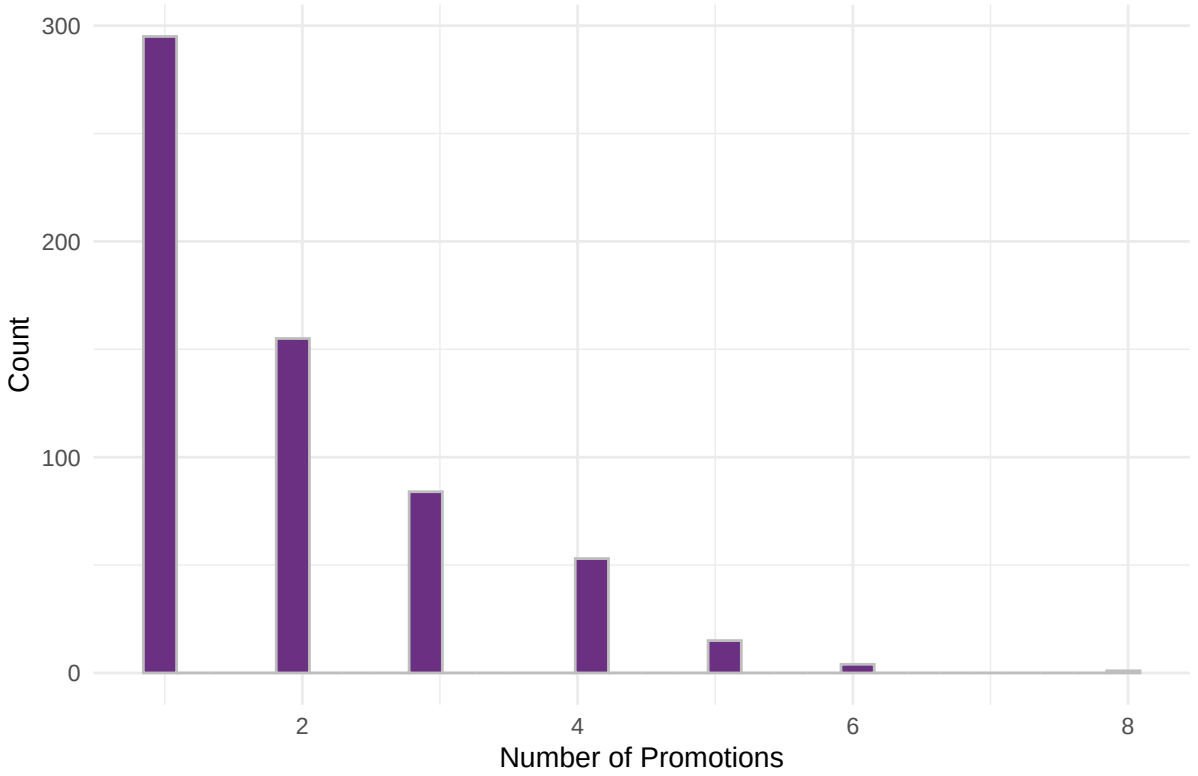

Figure 2: Distribution of number of promotions a current employee has received

We see from the above graph that the distribution of number of promotions given out to an employee is right-skewed, suggesting a Poisson distribution is reasonable, with count data often following such a distribution. We note that although many employees were not given any promotions, we will not be fitting a zero-inflated Poisson model because we do not consider these employees to be in a separate group nor do we consider them otherwise unable to receive a promotion.

We next calculated the mean productivity and mean leadership scores of each employee over all quarters in which they worked. To do this, we first had to convert each leadership description to a numeric value. We decided upon 0 for "Needs improvement," 1 for "Appropriate for level," and 2 for "Exceeds expectations" for simplicity. This was done so we have a single value of each score for every employee, which is needed for the model we are about to fit in the next section. We then added new variables for the number of quarters worked and the number of promotions received for each employee. Finally we created a new condensed tibble where each row corresponds to one employee, which will soon be needed because we are going to model a count for each employee.

Methods

Because we want to model the influence of gender while controlling for other relevant variables on the number of promotions an employee has received during their time at Black Saber, which is in essence a count, we choose to fit a generalized linear model with a Poisson response. It is difficult to assess the assumption of linearity of log(number of promotions received) because we have many discrete explanatory variables.

Our model included gender, team, role seniority when the employee first started (we checked that this is not the same for everyone), mean productivity score over all quarters worked, and mean leadership score over all quarters worked as explanatory variables. We added an offset for the number of quarters an employee has worked because another core Poisson assumption is that the response variable be a count per some unit of time or space. This unit is usually fixed, but can be different for each observation if we introduce it into the model as an offset. Because we model the log of the mean in Poisson regression, we added log(number of quarters worked) as our offset so our predicted values are scaled appropriately. We saw that our dispersion parameter for this model is taken to be 1, suggesting no evidence of overdispersion. We also ran a goodness of fit test and saw that if we had a true model, the probability of observing the deviance 142.7349115 on 587 degrees of freedom is practically 1, suggesting no evidence of lack of fit.

Results

The coefficients for the model are given below in Table 2. We removed the team coefficients because they were numerous and not very relevant to our discussion. We can interpret the coefficient on the gender explanatory variables by exponentiating: eˆ-0.0712=0.931 and eˆ-0.0868=0.917, indicating that both women and employees who preferred not to state their gender received about 10% fewer promotions than men on average, controlling for mean productivity, leadership scores, and seniority level when started and offsetting for amount of time worked. The P-value for both groups however are not statistically significant.

Table 2: Promotion Model Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-1.3524854	0.6571727	-2.0580365	0.0395866
Gender: Women	-0.0712409	0.0789016	-0.9029074	0.3665750
Gender: Prefer not to say	-0.0867826	0.2393505	-0.3625754	0.7169221
Mean Productivity Score	0.0026908	0.0032311	0.8327701	0.4049744

	Estimate	Std. Error	z value	Pr(>|z|)
Mean Leadership Score	-0.0869104	0.3554510	-0.2445074	0.8068378
Role Seniority when Started	-0.3949200	0.0810723	-4.8712091	0.0000011

Does the gender of the employee have an unfair impact on the salary they receive?

Because we used the same data as the previous research question for this research question, we will skip the data exploration and wrangling discussion.

Methods

We want to model the influence of gender on salary while controlling for other relevant variables. We use a linear mixed model because we want to use the salary information for all employees across all quarters, but we evidently cannot treat these observations as independent because one's salary in a particular quarter is very likely to be dependent on their salary in a previous quarter. However if we treat the employees themselves as a random effect, we can satisfy the independence assumption for linear mixed models because we have no reason not to treat the data as independent across employees. Thus we fit a linear mixed model with salary as the response variable and gender, productivity score, leadership level, role seniority, and team as explanatory variables, with employee ID as a random intercept.

Figure 3: Residual Plot for Salary Model

We see from the residual plot that conditioned on our chosen explanatory variables, the residuals are fairly well dispersed about the 0 line, with there being a roughly equal number on each side

and a greater concentration of points closer to the line. We conclude that a linear Gaussian response is reasonable, but the variances do not seem to be constant, with greater dispersion occurring at the middle than at the ends. We also have relatively fewer data points that correspond to high salaries, but we know we don't have missing data, so this is likely just due to the nature of how salaries are usually distributed within a company. We keep in mind that our nonconstant variance assumption may be violated, but linear mixed models are quite robust to some assumption violations so we proceed with our model.

Results

Our coefficient for women in our model indicates that in a given financial quarter, women at Black Saber have been paid an annual salary that is $2239 less on average than that of the men with a confidence interval of (-2776.366320, -1700.993037), controlling for the other aforementioned predictors. Because this interval does not contain 0, we can be 95% confident that female employees receive lower wages independent of the other predictors. Our model also indicates that employees who preferred not to state their gender have been paid an annual salary that is $1150 less on average that that of the men with a confidence interval of (-3160.411236, 859.836852), controlling for the same predictors. We do not take this as evidence of anything because the confidence interval contains 0.

Discussion & Conclusions

We fit a number of models in an effort to answer our research questions and obtained insignificant P-values for almost all of them. Thus we conclude by stating that we have found no evidence of an undue influence of gender on the AI system that conducts the initial hiring process, the interview scores in the third stage of hiring, or the number of promotions given to an employee. However, we are 95% confident that the mean annual salary of female employees is between \$1700.99 and \$2776.37 less than that of their male counterparts, independent of the other predictors. We must emphasize that stating we did not find any evidence of gender discrimination in a particular process does not imply we found evidence indicating there was none. More can be done to explore this; we give some suggestions for this below.

Strengths & Limitations

We believe we chose viable models for most, if not all of our research questions, as we made appropriate choices based on the response variable. We also made sensible choices about which predictors to use and transformed them when needed so they could be used in our models. Finally we made sure to check model assumptions, so we feel we have done all we can with the data we were given.

Along with possible violations of model assumptions that have already been discussed in the previous sections, we now discuss possible limitations of our analysis.

We were tasked with examining whether there was latent and unfair bias in the hiring and internal practices of Black Saber, but gender was the only information that was a potential source of discrimination that we were granted access to. It's also important to investigate whether there is discrimination on the basis of other protected or minority classes, such as those pertaining to race or ethnicity. We were told that an EDI initiative is being considered at Black Saber and we encourage such an initiative to collect data on other classes applicants and employees may belong to so that a similar analysis to the one done here can be done in the future with that data.

We also note that the employee data we were sent only consists of current employees. This is a potential limitation of our last two models because it could imply that we don't have data that is representative of Black Saber's true salary and promotion practices. For example it could be the case that women and those who preferred not to state their gender were frequently given lower wages and less opportunities for advancement and so they left the company. We were told that the company has a high retention rate, but we cannot verify this or if it applies equally across all gender groups with the data we currently have. We suggest that data on an employee be kept even if they leave the company, ensuring that a potential future analysis will have access to less biased data.