Association for Information Systems

AIS Electronic Library (AISeL)

Wirtschaftsinformatik 2023 Proceedings

Wirtschaftsinformatik

10-9-2023

A Longitudinal Examination of AI Fairness on Online Labor Markets

Brittany Green
University of Louisville, United States of America, brittany.green.1@louisville.edu

Manju Ahuja
University of Louisville, United States of America, Manju.ahuja@louisville.edu

Rui Sundrup
University of Louisville, United States of America, rui.sundrup@louisville.edu

Ryan Quinn
University of Louisville, United States of America, ryan.quinn@louisville.edu

Follow this and additional works at: https://aisel.aisnet.org/wi2023

Recommended Citation

Green, Brittany; Ahuja, Manju; Sundrup, Rui; and Quinn, Ryan, "A Longitudinal Examination of AI Fairness on Online Labor Markets" (2023). *Wirtschaftsinformatik 2023 Proceedings*. 81.
https://aisel.aisnet.org/wi2023/81

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A Longitudinal Examination of AI Fairness on Online Labor Markets

Research in Progress

Brittany Green[1], Manju Ahuja[1], Rui Sundrup[1], and Ryan Quinn[2]

[1] University of Louisville, Information Systems, Analytics & Operations, Louisville, U.S.A.
{brittany.green.1, manju.ahuja, rui.sundrup}@louisville.edu
[2] University of Louisville, Management & Entrepreneurship, Louisville, U.S.A.
ryan.quinn@louisville.edu

Abstract. While online labor markets (OLMs) provide many benefits including flexibility and data driven AI matching systems, gender and other social biases have been shown in OLMs, and research demonstrates AI can also perpetuate bias. However, previous OLM research assumes bias is static over time and independent of the AI algorithm. To help design OLMs that minimize the detrimental impact of biases on marginalized social groups, we investigate the interaction among individual characteristics and AI sources of biases over the long-term and evaluate auditing strategies using an agent-based simulation model. We also empirically investigate hiring bias using a cross section of data from a popular online labor market, and we use this data to inform our simulation. We then plan to develop and empirically test a framework to evaluate AI fairness and the interaction of different biases on OLMs and test an audit strategy to mitigate biases. We plan to extend the literature on OLMs by integrating fairness and intersectionality research to evaluate the impact of biases.

Keywords: *AI biases; AI fairness; Auditing*

1 Introduction

Previous empirical research on online labor markets has demonstrated the presence of race-based biases (Bertrand and Mullainathan, 2004; Ge et al., 2016), gender disparities (Chan and Wang, 2017; Bertrand and Duflo, 2017; Goldin and Rouse, 2000), as well as geographic bias and discrimination (Liang et al., 2018). Although general evidence of such biases in OLMs exists, developing a nuanced understanding of the interaction among individual characteristics and AI sources of biases could provide guidance for designing OLMs that minimize the detrimental impact of such biases on the income and career prospects of marginalized social groups. Given this, our first objective is to examine how a freelancer's identities impact hiring outcomes in OLMs. Further, the relative contribution of AI and individual characteristics, and their interaction on biases in OLMs is not well understood. Developing this understanding could help develop audits for biases in key spots in the recommendation process.

Moreover, the long-term impact of biases can be different from a short-term or one-time decision (Sun, Nasraoui and Shafto, 2020). Because freelancers' income and careers may be affected by cumulative effects of these decisions, it is imperative to understand the effectiveness and when to deploy interventions that may help mitigate biases. Thus, the second objective of the study is to examine the longitudinal effects of these biases. Finally, we seek to empirically examine the role of an important mitigation strategy— algorithmic audits—in managing OLM bias in the short- and long- term. This bias mitigation strategy has been conceptualized to validate fairness of an AI system by inspecting the data sources and logic throughout the lifecycle of an AI system (Robert *et al.*, 2020). We focus on two approaches: audits to establish fairness for large groups such as gender or race and those that account for intersectional fairness. The longitudinal effectiveness of this mitigation strategy has not yet been adequately tested.

Understanding where and when these biases occur (and on which groups) and how they interact with each other, can help design OLMs that incorporate audits, interventions, and mitigation strategies and therefore, produce more fair results. For example, a mitigation approach may reduce social biases, but if the AI matching systems continue to be based on data from previously biased decisions (made by humans), the impact of the mitigation approach can be limited. Furthermore, as AI systems learn (e.g., deep learning) the biases can become further magnified and calcified over time.

The theoretical foundations of this work are based on Organizational Justice Theory (OJT) and its two key components that shape individuals' perceptions in regard to fairness— procedural fairness (i.e., fairness in the process used to make decisions) (Leventhal, Karuza and Fry, 1980) and distributive fairness (i.e., fairness of decision outcomes) (Alexander and Ruderman, 1987). In our context, procedural fairness frames our objectives to investigate biases in OLMs and the development and testing of audit-based techniques aimed at reducing biases in the system. Distributive fairness refers to the fair allocation of jobs and pay for freelancers over time. However, the impact of the intersections of different biases can cause harm that is essentially ignored by broad descriptions of fairness.

We integrate OJT with Intersectionality Theory, which suggests that people uniquely experience biases based on their intersectional identities or exposure to multiple types of bias (Crenshaw, 1989). We consider distributive fairness of career outcomes and pay differences from the perspective of the interactions of biases from intersecting identities. We use this integrated lens to longitudinally compare the performance of an audit system for large groups versus those with a particular focus on intersectional experiences.

To investigate these questions, we are building an agent-based simulation model (ABM) that simulates an OLM with an AI matching algorithm based on data. As with any model, an ABM is a simplification of reality and will not address every variable or mechanism that will occur in real life. However, one advantage of an ABM is that it enables us to identify minimal necessary conditions for observed outcomes. Minimal necessary conditions are conditions that have high leverage in a system, and are therefore useful for directing our attention when examining potential interventions. Additionally, the ABM methodology provides the ability to investigate many confounding factors in a long-term experiment—something usually infeasible and too expensive to implement in longitudinal lab or field settings (Zhang et al., 2020). To elicit freelancer and job characteristics, we seed the ABM using data from a real-world system including a subset of scraped freelancer and job profiles from an OLM..

Our work may contribute to the emerging literature on social justice in digital technologies in significant ways.We seek to not only provide a snapshot of a current OLM, but also we seek to develop a deeper understanding of the longitudinal dynamics of biases in OLMs. Further, we seek to empirically examine the role of an important mitigation strategy— algorithmic audits—in managing OLM bias in the short- and long- term. In particular, we focus on two approaches: audits to establish fairness for large groups such as gender or race and those that account for intersectional fairness. Audits are a bias mitigation strategy that has been conceptualized to validate fairness of an AI system by inspecting the data sources and logic throughout the lifecycle of an AI system (Robert et al., 2020). The longitudinal effectiveness of this mitigation strategy has not yet been adequately tested. We extend the literature on OLMs by integrating fairness and intersectionality research to evaluate the impact of AI biases. This study is aimed at providing a theory of long-term effects of biases within OLMs for scholars and guiding practice for users, and developers of OLMs.

2 Preliminary Analysis

To investigate sources of bias on an online labor market, we use data collected from a major online labor market. This data contains most information from the profiles of software developer, website, and mobile app development jobs and the corresponding profiles of freelancers that applied to these jobs. We then merged the freelancer profile information with the job information by freelancer applicant. In our preliminary analysis, we do not consider all variables, since many variables are not appropriately formatted for analysis (e.g., free text variables).

In our data, some freelancers left the platform after applying to a past job, and their profiles were not available to be captured. Due to these missing freelancer profiles, only jobs with less than 15% missing freelancer profiles were kept in the dataset. If the hired freelancer was missing from the data, all applicants related to this job were also dropped from the analysis, as we are interested in hiring outcomes. A limitation of our data due to the online labor market website layout is that only 20 or fewer randomly selected freelancer applicants and the hired freelancer were available for each job, even if more freelancers applied to this job. Due to this, we currently focus on jobs with 19 applicants or fewer. In total there were 32,187 jobs that have 19 applicants or fewer. After removing jobs with more than 15% missing, we have 24,373 jobs. Because our analysis focuses on gender, we only keep jobs that have both a perceived man and a perceived woman in the applicant pool. We also only consider jobs where at least two freelancers apply. Within this data, some freelancers did not post a bid amount, and we remove these freelancer applicants' observations for this preliminary analysis. After these steps, the data contain a total of 13,697 jobs. In our preliminary analysis, we only consider the variables in Table 1. These variables are a subset of freelancer information from job profiles, since we attempt to limit bias from job differences through our modeling approach.

Gender is not directly visible on the platform, but employers likely infer gender through pictures or first names. To emulate employer hiring, we hire individuals from Amazon Mechanical Turk (AMT) to determine a freelancer's perceived gender from freelancer pictures. The gender options available are a man, a woman, no people, two or more people, and no picture categories. Two individuals from AMT labeled each picture, and agreement between these two individuals determined the freelancer's perceived gender. If there was not agreement, another round of AMT labeling was conducted, and a member of the research team labeled the freelancer as well. The subsequent majority vote from all labellers was used to determine a freelancer's perceived gender. When a picture was labeled as hard to tell gender, no people, no picture, or two or more people, we instead use the results from an AI based gender labeler that determines perceived gender based on first names. We acknowledge the simplification of perceived gender in this study to a man and a woman only.

2.1 Preliminary Empirical Analysis of Hiring Bias

To investigate hiring bias, we fit a conditional logistic regression model where the dependent variable is hired or not and the independent variables include a variety of freelancer characteristics visible in the freelancer's profile as shown in Table 1 (McFadden, 1973). The model is stratified by job to negate characteristics that may not be observed between jobs (McFadden, 1973). In this preliminary analysis, we use a subset of variables that have been cleaned for the analysis. We also include job fixed effects.

Table 1 shows the log odds of hiring where perceived women are less likely to be hired. The log odds ratio coefficient estimate of about -0.33 suggests that a perceived woman's odds of being hired are 28.01% lower than a perceived man's odds of being hired. Additionally, when a freelancer and employer are in the same country, the freelancer is more likely to be hired. Here the odds of being hired are 30.90% more for a freelancer in the same country as the employer compared to the odds of a freelancer living in a different country than the employer. Consistent with expectations, the higher the bid amount, the less likely the freelancer will be hired, and the higher a freelancer's rating, the more likely the freelancer will be hired. Unexpectedly, the more educational degrees a freelancer has, then the less likely they are to be hired. We are still examining this trend.

Table 1. Model Results

Variable	log(OR)	95% CI	p-value
Freelancer rating	0.22	0.21, 0.23	<0.001
log(Bid amount)	-0.79	-0.83, -0.75	<0.001
Perceived gender - a woman	-0.33	-0.38, -0.27	<0.001
Freelancer description present	0.70	0.49, 0.90	<0.001
Total days on platform	0.00	0.00, 0.00	<0.001
Education	-0.08	-0.10, -0.05	<0.001
Number of certifications	0.00	-0.01, 0.01	>0.9
Employer-Freelancer country match	0.27	0.18, 0.36	<0.001

2.2 Agent-based Simulation Model Setup

Our second objective is to investigate the longitudinal impact of employer bias in combination with an AI ranking system on freelancer careers. We do this in an agent based simulation model by modeling social bias through ratings bias and hiring bias towards freelancers with certain identities, partially based on what we learned in the preliminary empirical analysis. We model AI bias by using data from past hires containing social bias as the input to an AI system that ranks freelancer applicants for new jobs.

Given the page limit of this article, we provide a condensed explanation of the agent-based model and validation steps. The ABM iterates through six steps: 1) employers post jobs, 2) freelancers apply to jobs, 3) once a predetermined length of time after a job is posted, an AI system ranks applicants for this job, 4) If auditing is turned on, the system will re-rank applicants based on a given fairness metric, 5) the employer hires based on a combination of AI ranking, freelancer characteristics, and potential bias, and 6) employer rates freelancer performance on completed job. These steps enable us to examine human and AI bias because step 5 can capture human bias and step 3 may potentially perpetuate, create, or intensify any bias based on previous hiring data characteristics or imbalances. The iteration of these steps also allows us to simulate AI learning over time in conjunction with applicant and employer interactions, so that we can examine longitudinal effects.

We are currently in the validation phase of the simulation including but not limited to conducting T-tests to investigate the difference between identity groups and validating the results of our simulation with empirical data from an online labor market as discussed in Section 2.1. After validation, we will conduct multiple experiments to investigate the longitudinal impact of employer bias towards multiple freelancer identities.

3 Next Steps

We focus on two main aspects for future work. First, we plan to do a more comprehensive empirical analysis and implement further simulation experiments manipulating aspects of social and AI bias and platform characteristics. Second, using the simulation model, we will investigate the longitudinal impact of mitigation strategies in managing OLM bias in the short- and long- term, with a focus on auditing. Auditing investigates if the AI system in an OLM perpetuates bias when ranking freelancer applicants for employers and then implements an appropriate mitigation strategy. Using the simulation model, we plan to compare audits that deliver fairness to groups focusing on single marginalized identities to audits that deliver fairness to people with intersecting marginalized identities. In step 4 of the simulation, we will either audit the system based on fairness defined for all intersecting marginalized identities or fairness defined for larger marginalized identity groups. We then examine the impact of both audit options by comparing the career outcomes of all identity groups from both audit options. Our goal is to forward a contingency theory of audit mechanisms and their effectiveness in different situations in addressing social and AI biases, to provide more fair outcomes.

References

Alexander, S. and Ruderman, M. (1987) ‘The role of procedural and distributive justice in organizational behavior’, *Social justice research*, 1(2), pp. 177–198.

Crenshaw, K. (1989) ‘Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics’, *The University of Chicago legal forum*, p. 139.

Leventhal, G.S., Karuza, J. and Fry, W.R. (1980) ‘Beyond fairness: A theory of allocation preferences’, *Justice and social interaction*, 3(1), pp. 167–218.

McFadden, D. (1973) *Conditional logit analysis of qualitative choice behavior*. Available at: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf (Accessed: 7 March 2023).

Robert, L.P. *et al.* (2020) ‘Designing fair AI for managing employees in organizations: a review, critique, and design agenda’, *Human–Computer Interaction*, 35(5-6), pp. 545–575.

Sun, W., Nasraoui, O. and Shafto, P. (2020) ‘Evolution and impact of bias in human and machine learning algorithm interaction’, *PloS one*, 15(8), p. e0235502.

Zhang, J. *et al.* (2020) ‘Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework’, *Information Systems Research*, 31(1), pp. 76–101.