Gender representation and the adoption of hiring algorithms: Evidence from MBA students and executives*

Patryk Perkowski

Columbia Business School

Job Market Paper

Click here for the latest version

November 20, 2022

Abstract

I examine how job performance and diversity considerations shape recommendations for adopting algorithms in hiring. Between 2019 and 2022, around 400 business managers and executives coded up and evaluated algorithms aimed at improving hiring at a firm. Although these algorithms would lead to large performance improvements on average if implemented, managers were unlikely to recommend adoption if they decreased the number of female hires at the organization. Algorithms that decreased the number of female hires were half as likely to be adopted as those that had no impact or increased it. These results hold while controlling for the algorithm's impact on job performance, and are not present for two other protected classes, age and country of origin. I rule out that this behavior was driven by fear of illegal discrimination. Instead, using a regression discontinuity design, I find evidence that this behavior is consistent with managers exhibiting "female representation loss aversion", whereby even small decreases in the number of women at the firm lead managers to reject hiring algorithms. I explore what modeling strategies or programmer characteristics can mitigate the likelihood of decreasing female representation, and find that programming approaches such as those that excluded protected demographic predictors or considered multiple models were less likely to decrease the representation of women. I also provide evidence of a complementarity between firm screening practices and the returns to algorithmic decision-making. I show that the performance impacts of algorithmic hiring are highest for firms that screen for technical workers with machine learning skills, illustrating an important complementarity between technical hiring and hiring for machine learning skills. I conclude by discussing the implications of this paper for our understanding of algorithmic aversion, algorithmic bias, and the returns to algorithmic decision-making in business.

Keywords: hiring algorithms, algorithmic aversion, gender, algorithmic decision-making
JEL Classification: M5, J7, M15

*Patryk Perkowski (PPerkowski22@gsb.columbia.edu). I thank Erica Bailey, James T. Carter, Bo Cowgill, Fabrizio Dell'Acqua, Bruce Kogut, Jean Oh, Katherine Sun, and Jordan Lionell Weatherwax for helpful feedback. I thank Nicole Hawro for excellent research assistance, and Angela Ryu for help accessing the data. All errors are my own.

1 Introduction

Rather than rely on human intuition, firms are increasingly delegating decision-making authority to algorithms. Rapid changes in data collection, storage, and processing technologies have led to the rise of data-driven decision making in businesses (Brynjolfsson et al. 2003; Brynjolfsson and McElheran 2016a,b), whereby firms rely less on human intuition and more on data. In some instances, firms have moved from using data to influence decision making to completely delegating decision-making authority to algorithms. Evidence from the past decade indicates that algorithms can outperform humans in settings as diverse as chess (Hassabis 2017), medical diagnoses (Rajpurkar et al. 2018), and speech detection (Assael et al. 2016). Such developments have coincided with algorithmic adoption in business domains such as service operations, product development, and marketing and sales. Overall, the rise of algorithms in business will have fundamental impacts on how businesses compete and operate.

At first blush, algorithms appear prime for adoption in the field of recruitment. First, hiring is inherently a prediction problem, whereby a firm tries to forecast a given applicant's on-the-job performance, usually using information obtained from the application process such as a resume or interview.[1] Given that recent developments in machine learning and artificial intelligence are advances in statistical prediction (Agrawal et al. 2019), hiring seems well suited to the use of algorithms. Indeed, recent empirical evidence suggests that algorithms are more effective at identifying top job performers than humans (Kuncel et al. 2014; Chalfin et al. 2016; Cowgill 2020; Li et al. 2021). Second, the rise of human capital management platforms, such as worker and applicant tracking systems, has increased the quality and quantity of data that firms have about their workers and job applicants (Aral et al. 2012). Increased data helps improve prediction accuracy, which would thus increase the returns to algorithmic hiring. Third, technological advancements have increased the returns to selective hiring, making it more important to have accurate predictions of potential hires' job performance (Cowgill and Perkowski 2022). Fourth and finally, the rise of job search platforms like Indeed and LinkedIn has increased application volumes, thereby increasing firm screening costs (Cappeli 2001); algorithmic approaches offer a solution with lower marginal costs than human screening. Overall, improvements in prediction technology, increased data availability, and higher firm screening costs favor algorithms playing a prominent role in hiring.

[1] Firms may also try to forecast other information about a candidate, such as their desired salary. This is also inherently a prediction problem for the firm.

Still, firm adoption of algorithms in hiring has lagged behind algorithmic adoption in other business functions. McKinsey's Global Survey of AI in 2021, for example, found that while over half of respondents reported using analytics in at least one business function, only eight percent reported using it in talent management.[2] Firms were around three times as likely to use algorithms for product and service development, or service operations, and twice as likely to use it for customer service analytics than talent analytics.[3] These survey results suggest that despite the benefits to algorithms in hiring, there are meaningful barriers to adoption.

One reason why adoption lags in hiring is that it presents statistical challenges that other business settings do not. Perhaps the most salient of these challenges is the difficulty of obtaining an outcome measure for the algorithm to maximize (Tambe et al. 2019). Job performance is notoriously difficult to measure (Levinson 2003), and managers may seek to optimize outcomes other than job-performance (such as retention, salary requirements, or a combination of these measures). This increases the difficulty of deploying algorithms in hiring, compared to other business domains where the outcome measure is more easily measurable (for example, ad clicks). Additionally, analytics done in the personnel arena suffer from issues regarding smaller sample sizes and significant sample selection issues.[4] These obstacles increase the difficulty of prediction, are less pressing in other business applications, and depress the usefulness of algorithms in hiring. Overall, these statistical challenges create significant obstacles to algorithmic adoption in hiring.

In addition, algorithmic hiring raises ethical concerns regarding bias, discrimination, and distributional outcomes, which are paramount to business and society. Articles in scholarly journals and in the public press document many instances of algorithmic bias, whereby algorithmic approaches either codify or introduce new bias against under-represented candidates.[5] These concerns are front and center for policymakers, firms, and employees, alike. In 2019, United States Senator Ron Wyden introduced the Algorithmic Accountability Act, which would require that companies perform audits of

[2]This percent includes firms using algorithms for recruiting and for retention, so the percent using algorithms in hiring is almost certainly lower. See https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021 for more information.

[3]This may reflect the results in Bhatia and Meier (2022), who find that executives show a bias towards customers rather than employees using data from earnings calls and board composition.

[4]For example, suppose that an organization trains an algorithm to maximize job performance. However, the organization does not have data on the job performance of individuals it does not hire. This is an important sample selection bias that the organization must deal with. This refers to the "selective labels problem" in computer science. For more information on the selective labels problem, see Kleinberg et al. (2017); Lakkaraju et al. (2017); Cowgill (2020); Rambachan and Roth (2020).

[5]See, for example, https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias.

their algorithmic systems and report the findings to the Federal Trade Commission.[6] In New York City, a new law will require that employers conduct independent audits of their automated tools to test for bias.[7] Meanwhile, firms have begun partnering with academics to audit their algorithmic approaches to hiring. The startup pymetrics, which offers a machine-learning pre-employment assessment, teamed up with a series of academic researchers to run audits to evaluate their system for bias (Wilson et al. 2021). The start-up HireVue also participated in such an audit by working with Richard N. Landers, a prominent Industrial-Organizational Psychologist.[8] Such concerns are also commonly held amongst the general population. A Pew Research survey, for example, found that 42 percent of American adults believe that hiring algorithms would do a worse job than humans when it comes to hiring candidates from diverse backgrounds (compared to 27 percent who thought they would do better) and 58 percent belief algorithms would do worse than humans at evaluating candidates with non-traditional work experience.[9] In sum, business and society are skeptical that algorithms can provide unbiased predictions. And, these concerns have compelled firms and policymakers alike to pay special attention to the distributional consequences of hiring algorithms across racial, gender, and other demographic lines.

To be sure, the rise of algorithms has increased the codifiability and testability of the impact on hiring policies on demographic representation. Algorithms have given decisionmakers the ability to carefully measure the impacts of the algorithm and any distributional shifts in the diversity of hires.[10] Even black-box algorithms, which obscure how inputs are combined, can be audited and tested for demographic impacts.[11] The use of these tools will require that executives decide between algorithms of varying effects (for example, choosing to implement an algorithm that will increase the female representation but decrease the number of workers above the age of 50 by six percent distribution, versus one that has the opposite effect). Understanding how these tradeoffs influence algorithmic adoption decisions has important consequences for our understanding of hiring, human resource strategy, and worker careers more generally.

[6] See https://www.congress.gov/bill/116th-congress/house-bill/2231 for the text of the bill.

[7] https://news.bloomberglaw.com/daily-labor-report/new-york-city-ai-bias-law-charts-new-territory-for-employers

[8] For more information, see https://www.hirevue.com/press-release/independent-audit-affirms-the-scientific-foundation-of-hirevue-

[9] See https://www.pewresearch.org/internet/2017/10/04/americans-attitudes-toward-hiring-algorithms/.

[10] It is interesting to note that several papers find that algorithms can improve diversity, relative to human decision makers who are more biased. For example, an number of papers, including Cowgill (2020), Li et al. (2021), Pisanelli (2022b) and Pisanelli (2022a), present evidence that the use of algorithms in hiring increased the number of hires that were female or were racial minorities compared to the status quo approach of humans. However, the extent to which this is true in broader settings is unclear.

[11] See, for example, https://www2.deloitte.com/us/en/pages/advisory/articles/black-box-artificial-intelligence.html.

I examine these tensions using four years of data from an assignment in a People Analytics course at a top business school in the United States. From 2019 to 2022, over 450 MBA and Executive MBA (EMBA) students (collectively referred to as "managers") were tasked with writing hiring algorithms in order to improve hiring practices at an anonymous (and fictitious) company (referred to as "Firm F." The managers received data concerning approximately 2,000 workers, 689 of whom were currently employed by the company, and 1,258 of whom were not currently employed but were eligible to be hired. Managers had access to workers' demographic characteristics (such as gender, age, and education), employment characteristics (such as work status), and current employer characteristics (such as firm size and workplace flexibility). For the 689 hired workers, the managers had a measure of job performance that was unbiased and without noise. This data represents data that is collected by human capital management systems such as ADP and Workday, albeit with a smaller sample and fewer covariates.

The assignment asked managers to analyze this data and develop an algorithmic proposal to improve hiring practices at this firm through a series of five steps. First, the managers were asked to determine the predictors of being hired by the firm, and of job performance at the firm. Second, the managers were asked to formulate hypotheses on how the firm could improve its hiring practices using the results of their analysis. Third, the managers quantified and translated their hypotheses into hiring algorithms that they then used to select (i) 20 workers in the applicant pool whom the firm should have hired, and (ii) 20 workers who were currently employed by the firm but who should never have been hired.[12] Fourth, after making their hiring recommendations, the managers were given performance data for all 1,258 potential workers and had to evaluate the proposals' impact on job performance and workplace diversity. Finally, managers asked whether they would recommend that the firm adopt their proposals or maintain the status quo of hiring by humans. The exercise was designed to resemble a typical task conducted by a People Analytics team or by human resource management consultants, with a special focus on relating the adoption recommendation to the performance and demographic representation impacts.[13]

[12]In theory, the use of algorithms in hiring may also change the number of employees who are hired (for example, if the algorithm selects workers who have lower salary requirements or have a higher willingness to supply for the firm. In this setting, I hold constant the number of employees influenced by the change. However, it is possible to simulate the impact of the algorithm across the size threshold. I will explore this in future work.

[13]The task was designed in a way to limit many of the statistical issues mentioned above. The setting features a job-performance measure that is unbiased, and managers were instructed to ignore other outcome variables such as salary demands. The aim was to focus attention on relating the adoption recommendation and the performance and demographic representation impacts, and not the other challenges that would depress adoption recommendations.

There are a number of reasons why this task presents a natural setting to study how the demographic representation impacts of algorithms shape their adoption decisions. First, the setting allows me to hold constant other factors that may impact algorithmic adoption (such as the amount of training data or the existence of a measurable job performance outcome) and thereby to zero in on how demographic representation shapes algorithmic adoption. Second, the setting provides a unique measure of algorithmic adoption. It is notoriously difficult to get data on firm technology adoption decisions; my setting contains a measure of algorithmic adoption recommendations, which I can then relate to the performance outcomes of algorithms. Third, the setting allows me to bypass the Fundamental Problem of Causal Inference and estimate the causal impact of algorithms on firm outcomes. Understanding how the effects of an algorithm impact adoption rates requires an unbiased measure of how the algorithm changes job performance and demographic representation. While it is possible to estimate how a given algorithm would influence demographic diversity (for example, by using algorithmic audits as described earlier), such audits are more difficult for job performance since this measure is only collected conditional on having been hired. Although no proposal was ever implemented, I can assuage this concern using a simulation method described in Section 3.2 that uses the forty workers whose employment status is changed by the adoption of algorithms. Overall, the task presents a creative setting that bypasses many of the difficulties in studying algorithmic adoption decisions.

My analysis proceeds in the following steps. First, I estimate the causal impact of adopting a hiring algorithm on Firm F's outcomes by comparing outcomes for the forty candidates whose hiring status changes as a result of the algorithm. I am particularly interested in the impacts on job performance and the demographic distribution of hires, both on average and across the distribution of managers.[14] Second, I estimate how these impacts are related to the adoption recommendation given by the manager. I regress the adoption recommendation on the performance and demographic impacts of the algorithm to see what is the best predictor of adopting. Third, I investigate whether there is support for various mechanisms that could be driving the adoption decisions I observe, such as fear of illegal discrimination. I also conduct two supplementary analyses to further elucidate behavior in my setting. Both involve relating the impacts of the algorithm (on job performance and on the demographic distribution of hires) to the demographic identity of the manager who wrote it, and to the algorithmic methods used. For example, some managers used advanced machine learning techniques to come

[14] In a sense, this second analysis mirrors recent papers that give a team of analysts the same data and study the distribution of conclusions that the group comes to. See, for example, Silberzahn et al. (2018)'s study of bias in soccer.

up with predictions of job performance, while others came to their recommendations using summary statistics that would be typically available in a People Analytics dashboard. I estimate how these modeling choices and the manager's background influence the impacts of the algorithm. This allows me to see whether it is possible to predict the performance or demographic impacts of an algorithm from its coder and modeling strategy

My results indicate that if implemented, the hiring algorithms would have accepted candidates with much higher job performance scores than the status quo. The 20 recommended hires had a job performance score that was on average around 470 points higher than the twenty that the algorithms suggested to fire. The 20 recommended fired candidates had a job performance score of 2,345, indicating that the algorithm would lead to a 20 percent improvement in job performance of the firm's hires. This is true not just for the average candidate recommended by the algorithm, but also throughout the performance distribution of recommended candidates. Results from quantile regressions reveal that the algorithms increase the performance of the bottom 10 percentile of workers by 30 percent, and the top 10 percentile of workers by 15 percent. Moreover, over 80 percent of managers wrote a hiring algorithm that increased the job performance of hires, with an average increase of 32 percent. Those whose algorithms decreased job performance did so by 11 percent on average. Overall, the subject population was effective in writing algorithms that improved the quality of hires as measured by job performance.

Although the algorithms tended to increase job performance, the algorithms had more nuanced effects on the demographic distribution of hires. I generate binary indicators for candidates from three protected classes: being female, being age 40 or older, and being from Latin America, and examine how the algorithms impact these measures. My results indicate that the algorithms had negative impacts on the proportion of female, older, and Latin American hires, reducing their numbers by 13, 40, and 90 percent, respectively. Meanwhile, the hiring algorithms led to hires that skewed male, younger, and were more likely to be from Asia, North America, and Western Europe. While the average impact on the gender diversity of new hires is negative, this effect features some heterogeneity. Around half of managers increase the proportion of women, and they do so by 30 percent on average, while the other half decrease it, and they do so by 39 percent on average. Thus, while adopting the hiring algorithm would have increased job performance for the majority of managers, it would also lead to large changes in the demographics of hired workers.

How did these performance and diversity considerations shape adoption recommendations? I next relate each manager's adoption recommendation to the impact of their algorithm on measures of job performance and diversity. My results indicate that managerial adoption decisions are especially influenced by the algorithm's impact on gender diversity. Algorithms that reduce the number of female hires decrease adoption rates by 35 percentage points, or a 50-percent decline relative to the control mean of 70 percent. This holds true even controlling for the algorithm's impact on job performance, and on the other protected categories. Meanwhile, managers are less responsive to the impacts on other demographic characteristics. They show some skepticism in adopting algorithms that have an adverse impact related to age, but this effect is not statistically significant and it is less than half of the size of the gender effect. Meanwhile, adverse impacts on national origin have no impact on adoption decisions. These results highlight the unique role played by gender diversity considerations in shaping algorithmic adoption decisions.

I then examine several potential mechanisms for the decrease in adoption rates for algorithms with an adverse effect on female representation. My results indicate that fear of a discrimination lawsuit is not the primary driver of algorithmic adoption decisions in my context. According to this theory, managers are wary of adopting algorithms that lead to adverse impacts on protected categories, since it is illegal to discriminate on these bases. However, my results indicate that managers do not respond to adverse impacts in two other protected domains – age and region of origin – as they do with gender. For these reasons, it is unlikely that fear of illegal discrimination is the primary motive for adoption decisions, unless they are particularly fearful of adverse impact claims regarding gender. I also rule out this possibility by using the EEOC's Four-Fifths rule in hiring, which states that disparate impact exists in a selection rule if the selection rate for a group is less than four-fifths of the rate for the group with the highest selection rate.[15] For example, an algorithm that accepts ten percent of male applicants but only three percent of female applicants would violate the Four-Fifths rule. I code up the average selection rate for men and for women for each algorithm, and tag the ones that violate the rule. Almost 80 percent of algorithms failed the test, and 60% of these had an adverse impact for women. Managers, however, still suggested adoption in over 70 percent of the algorithms that featured an adverse impact for men, making it unlikely that avoiding illegal gender discrimination was the primary driver of managerial decision-making in this context.

Instead, I find that the lack of adoption is due to "female representation loss aversion", whereby

[15] For more information, see https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-u

managers have an aversion to implementing *any* hiring algorithm that leads to fewer female hires than the status quo. A large literature in behavioral economics documents the presence of loss aversion, or the cognitive bias whereby humans prefer avoiding losses to acquiring equivalent gains ("losses loom larger than gains") (Kahneman and Tversky 1979; Brown et al. 2022). In my context, loss aversion with regard to female representation occurs if managers are averse to implementing hiring algorithms that lead to even a small decrease in the number of female hires (regardless of whether they satisfy the Four-Fifths rule or not). Moreover, this loss aversion is asymmetric in that the effect is concentrated amongst hiring algorithms that decrease the number of female hires, and not male hires.

I test the "female representation loss aversion" hypothesis by using a regression discontinuity design that compares adoption recommendations for hiring algorithms that marginally decrease the number of female hires, versus those that marginally do not (i.e., they keep the gender distribution of the firm the same or marginally increase the number of female hires). I first show that hiring algorithms to the right versus left side of this cutoff exhibit no systematic differences in their impacts on job performance or demographic characteristics other than gender. However, algorithms to the left of this cutoff were much less likely to be recommended for adoption: managers were 33 percent less likely to recommend the adoption of hiring algorithms that marginally decreased the number of female hires. This effect size is also large: moving from a marginal increase to a marginal decrease in the number of female hires represents around 75 percent of the overall effect of lowering female representation that I document. I also compare adoption recommendations for algorithms that marginally decrease the number of female hires, versus those that marginally increase the number of female hires, versus lead to no change in the gender breakdown at the firm, to illustrate the asymmetric nature of adoption recommendations.

Because managerial adoption recommendations are strongly shaped by an aversion to decreases in female representation, I explore what modeling strategies or manager characteristics predict whether an algorithm will reduce the number of female hires. I find that algorithms that excluded protected demographic information were 20 percentage points less likely to lower female representation, suggesting that blinding the algorithm to sensitive information was effective in my setting. Algorithms that tested multiple hypotheses and approaches were also less likely to decrease female representation, though this effect is about half as small. Meanwhile, the effects of programmer/manager demographic variables were more muted. I find no evidence that the gender, technical background, or educational institution of the manager influenced whether their algorithm decreased female representation or not. The one

manager covariate that did have an impact was whether the manager was an MBA or EMBA student. EMBA students were 26 percent less likely to write an algorithm that decreased female representation, suggesting that workplace experience may play an important role in determining algorithmic outcomes. Overall, these results suggest that increasing algorithmic adoption in the domain of hiring will require examining how various modeling approaches can mitigate the risk of decreasing female representation.

In my penultimate section, I zoom out of the discussion on how gender representation concerns affect adoption recommendations, and consider the broader question of whether there are complementarities between firm hiring strategies and the returns to algorithms in decision-making. The set-up of the manager's task allows me to examine how various firm hiring strategies impact the performance effects of algorithmic decision-making. I focus on three types of firm hiring strategies: (i) selective hiring (whereby a firm only hires candidates who have attended top universities); (ii) technical hiring (whereby a firm only hires candidates who have a Bachelor of Science degree); and (iii) machine learning hiring (whereby a firm only hires candidates who have machine learning skills). I simulate a set of outcomes where each manager corresponds to a firm whose hiring practices led to that worker being hired. For example, a manager with a BS from a non-elite university and with machine learning skills would correspond to a firm that hires for technical degrees and machine learning skills but not elite universities. I then relate the job performance effect of the algorithm to these three strategies and their interactions. I find evidence of a complementarity between technical hiring and hiring for machine learning skills, whereby technical workers with machine learning skills outperform both technical workers without machine learning skills, and non-technical workers with machine learning skills. Meanwhile, selective screening does not affect the returns to algorithmic decision-making, neither individually nor in tandem with the other two hiring strategies. These results suggest that unlocking the returns to algorithmic decision-making will require complementary changes to firm screening practices that focus on hiring for technical workers with machine learning skills.

This paper proceeds as follows. Section 2 describes the background and related literature, and highlights this paper's contribution to the literature on algorithmic adoption, the literature on the causes and consequences of algorithmic bias, and the literature on the impact of algorithmic decision-making on firm performance. Section 3 describes the setting, including the task, participants, and data. Section 4 outlines the empirical strategy while section 5 displays the empirical results. Section 6 discusses the implications of the results for algorithmic adoption and for firm hiring strategies, and

section 7 concludes.

2 Background and related literature

This paper is related to three interconnected literatures: the literature on algorithmic adoption in hiring, the literature on the causes and consequences of algorithmic bias, and the literature on the impact of algorithmic decision-making on firm performance. I begin by providing some background information and describe how my paper contributes to our understanding of these literatures.

2.1 The adoption of algorithms in hiring

This paper concerns the adoption of algorithms in hiring. A 2020 survey of human-resource managers found that the proportion of human resource departments that used predictive analytics jumped four-fold from 2016 to 2020.[16] Part of this growth is likely driven by an increasingly large ecosystem of start-ups that offer algorithmic services in hiring, such as Eightfold AI, Beamery, and FindEm.[17] Such startups (and tools developed in-house) have allowed organizations to use algorithmic tools in hiring throughout the organizational hierarchy, including C-suite executives.[18] These adoption decisions are likely driven by the benefits that algorithms offer relative to human and non-algorithmic approaches regarding quality and cost.

A number of empirical papers provide evidence that algorithmic hiring leads to better quality hires than traditional human approaches. For example, Chalfin et al. (2016) illustrates how machine learning can improve predictions of worker productivity in police officer hiring and teacher promotion decisions. Cowgill (2020) shows that the use of algorithms helps improve performance on various productivity measures including the likelihood of passing an interview, the likelihood of accepting a job offer, and on-the-job productivity when employed. Li et al. (2021) illustrates how a hiring algorithm that values exploration improves the eventual hiring rates of candidates selected for an interview relative to the firm's existing practices. These results suggest that algorithmic hiring can lead to hired candidates that are more productive than those brought about through human approaches.

[16] https://www.mercer.com/content/dam/mercer/attachments/private/global-talent-trends-2020-report.pdf
[17] See Raghavan et al. (2020) for an analysis of the claims and practices of such startups with regard to bias.
[18] See, for example, https://hbr.org/2015/04/hiring-c-suite-executives-by-algorithm.

Despite the evidence that suggests that algorithms can improve hiring outcomes, there are meaningful barriers to algorithmic adoption. First and foremost, there is a growing literature on algorithmic aversion, which seeks to unpack why human decision makers often ignore algorithmic recommendations, even when those recommendations outperform humans.[19] This literature has argued that a variety of factors influence the presence and size of algorithmic aversion, including being more responsive to algorithmic errors (Dietvorst et al. 2015), perceived capabilities of the algorithm (Longoni et al. 2019; Hertz and Wiese 2019), having a desire for human coworkers (Dell'Acqua et al. 2022), task characteristics (Castelo et al. 2019; Hertz and Wiese 2019) and financial incentives and framing (Greiner et al. 2022).[20] This literature has documented the human tendency to distrust and avoid algorithmic recommendations, even if they outperform human ones. Concerns regarding algorithmic aversion are also larger in hiring than in other domains. For example, the Pew Research Center's survey on automation in everyday life found that almost 70 percent of US adults say the development of hiring algorithms makes them feel somewhat or very worried, compared to 54 percent for driverless cars, and 47 percent for robot caregivers for older adults.[21] For these reasons, algorithmic aversion likely depresses adoption rates in the domain of hiring.

In addition to general algorithmic aversion, there are also statistical barriers that impede algorithmic adoption in hiring, especially relative to other business domains. One barrier is obtaining a dependent variable for the algorithm to predict. While dependent variables in other business domains are clearer (for example, ad clicks or product buys in marketing), obtaining such a measure in the hiring context is difficult. In theory, an algorithm can seek to maximize the expected job performance of new hires, but measuring job performance is notoriously difficult (Levinson 2003; Tambe et al. 2019). Moreover, even if one can get a measure of job performance, managers likely have other concerns such as salary requirements and retention. An algorithm that maximizes the job performance of new hires may fail in an organization where retention concerns are important; instead, an algorithm that maximizes the product of job performance and expected retention may fare better. A second barrier is limited data, both with respect to the unit of analysis and observability of the outcome. While an advertising or marketing team may have daily or even hourly sales data for a product, job performance outcomes

[19] See for Burton et al. (2020) and Mahmud et al. (2022) for systematic literature reviews of algorithmic aversion in human-machine decision-making. See Glikson and Woolley (2020) for a review of the empirical literature on human trust in artificial intelligence.

[20] This is also a literature on algorithmic appreciation, whereby humans are more likely to follow recommendations given by an algorithm. See, for example, Logg et al. (2019).

[21] See https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/. Part of this is due to skepticism that algorithms could do better than humans in the hiring process. 60 percent think algorithms would do worse in hiring candidates who fit well with a company's culture.

(however measured) are measured in more aggregate time periods (for example, quarterly performance reviews for workers), which limits the predictive accuracy of algorithms. A third limitation is time lag; it takes a while to observe job performance outcomes, which makes the programming difficult.[22] These issues have made it difficult to adopt off-the-shelf machine learning and algorithmic approaches into the hiring domain.

In addition to these statistical barriers, there are also ethical and normative challenges to algorithmic adoption in hiring. Bias, discrimination, and demographic representation present pressing issues for algorithmic adoption in hiring. A Pew Research survey, for example, found that 42 percent of American adults believe that hiring algorithms would do a worse job than humans when it comes to hiring candidates from diverse backgrounds (compared to 27 percent who thought they would do better).[23] Moreover, 58% believe that algorithms are worse than humans at evaluating job applicants with non-traditional work experience. These results highlight that humans are skeptical that algorithms are better than humans at screening candidates from diverse backgrounds and non-traditional work experiences. There are also concerns about algorithmic bias, whereby algorithms codify existing biases. Many articles in the popular press document the existence of algorithmic bias. Indeed, Cowgill et al. (2020a) provides experimental evidence that providing business leaders with prompts about the unavoidable nature of algorithmic bias depresses algorithmic adoption intentions. Overall, algorithmic impacts on diversity and bias are front and center when confronting algorithmic adoption decisions.

In this paper, I examine how these considerations shape algorithmic adoption intentions. In the empirical exercise, managers evaluate the performance of their algorithms in terms of on-the-job performance and demographic representation. I study how the decision to use algorithms or maintain the status quo of human decision-making is related to these effects. Both job performance and diversity considerations, including tradeoffs between various protected categories like gender, age, and national origin, shape the perceived performance of hiring systems, and digitization makes the codifiability of these tradeoffs possible. I study how the presence of algorithms in the decision-making process shifts outcomes along these dimensions, and how these outcomes shape adoption decisions. I do so in an environment where many of the statistical issues described above are controlled for through the design of the task. Moreover, I show that diversity constraints are a unique contributor to avoiding algorithms in the domain of hiring, suggesting that "female representation loss aversion" is an important

[22] There are, of course, some empirical workarounds to these problems. For example, practitioners may use a surrogate index (Athey et al. 2019) to come up with the predicted value of a long-term outcome given short-term outcomes.

[23] See https://www.pewresearch.org/internet/2017/10/04/americans-attitudes-toward-hiring-algorithms/.

phenomenon shaping human beliefs and behavior regarding algorithms.

2.2 Algorithmic bias

My paper also contributes to a burgeoning literature on algorithmic bias, which uses tools from computer science, economics, and management to examine what leads algorithms to make unfair and systematic errors against certain groups (Kirkpatrick 2016; Kleinberg et al. 2018; Cowgill and Tucker 2020; Rambachan et al. 2020). Understanding what causes algorithmic bias. and potential solutions to mitigate these concerns, is of utmost importance. This literature generally considers three broad sources of algorithmic bias, each of which has a different solution for reducing and eliminating bias: (i) biased input data, (ii) biased modeling approaches, and (iii) biased programmers.

First, algorithmic bias can stem from biased input data (Rambachan and Roth 2020; Cowgill et al. 2020c; Cowgill and Tucker 2020; Choudhury et al. 2020. Algorithms require input data to make predictions; if this input data does not represent the broader population, the predictions from this algorithm may lead to biased predictions. Perspectives regarding biased input data are captured by the popular phrase "bias-in, bias-out" and can manifest themselves in various ways. One example is sample selection bias, whereby an algorithm is trained on a dataset that is unrepresentative of the population at hand. Another is the problem of selective labels, whereby observed outcomes (for example, job productivity) are only observed for a subset of the population, which can be influenced by bias from humans (Kleinberg et al. 2017; Lakkaraju et al. 2017; Cowgill 2020; Rambachan and Roth 2020). If algorithmic bias is due to biased input data, then two solutions have been proposed: (i) having humans impute labels or outcome variables in regions of high confidence (De-Arteaga et al. 2018); or (ii) using representative data (Cowgill et al. 2020c), though such solutions may be difficult in the context of hiring.[24]

Second, algorithmic bias can stem from the modeling approach that translates inputs into algorithmic predictions. While the approach described in the previous paragraph examines the observations available for prediction (and how they are related to outcomes), this one focuses on the variables that go into an algorithmic prediction, and how those variables are combined to lead to a recommendation

[24]For example, it is difficult to predict job performance, which may make imputation unreliable. Moreover, organizations are unlikely to experiment with their hiring processes (Oyer and Schaefer 2010), which makes gaining representation data more difficult. These issues make the biased input data problem more pronounced in the domain of hiring.

(see, for example, Pope and Sydnor 2011). Prior work has examined how the inclusion or exclusion of specific characteristics, including gender (Goldin and Rouse 2000), criminal history (Agan and Starr 2017; Doleac and Hansen 2020; Cullen et al. 2022), and salary history (Agan et al. 2021; Hansen and McNichols 2020), influences outcomes across demographic groups. Some proponents argue that excluding information about demographic characteristics in algorithms can decrease bias. For example, LinkedIn's algorithm excludes the name, age, gender, and race of individuals to avoid bias in its job matching algorithm.[25] In this domain, algorithmic bias can be mitigated by paying careful attention to the variables that are allowed to enter the prediction. Similarly, a line of thinking argues that the quantitative level of sophistication of programmers can curtail algorithmic bias. For example, machine learning tools allow for not only more flexible functional forms, but also for programmers to hold out and test their algorithms before implementation, which are difficult for humans to do non-algorithmically. For example, algorithms can handle correlations between predictors in a way that is difficult for human decision makers. Overall, these approaches pay careful attention to the modeling approach that translates inputs into outputs, in order to reduce algorithmic bias.

Finally, algorithmic bias may be due to bias in the sample of computer programmers. The technology sector in the United States is one of the least diverse[26] Bias here can arise because the population of computer programmers is not representative of the population of those who will be impacted by the algorithm (Cowgill et al. 2020c; Cowgill and Tucker 2020). White and male programmers may not pay enough attention to issues regarding bias and discrimination, and their algorithms may feature more algorithmic bias than algorithms written by non-white and female programmers. In such cases, improving the demographic representation of programmers should decrease algorithmic bias.

This paper contributes to this literature by investigating which of these hypotheses are responsible for the adverse gender effects observed in the algorithms in my setting. While I cannot test whether biased input data influences algorithmic bias because my setting features no variation in the sample data that is given, I differentiate between the "biased modeling strategy" versus "biased programmer" hypotheses above. I find support for the former. Algorithms that excluded protected demographic variables or tested multiple models were less likely to lead to an adverse impact on gender representation, suggesting that focusing on algorithmic approaches may be an effective way to curtail algorithmic

[25] See https://www.technologyreview.com/2021/06/23/1026825/linkedin-ai-bias-ziprecruiter-monster-artificial-intelligence/ for more information.

[26] See, for example, the EEOC's special report on diversity in high tech. https://www.eeoc.gov/special-report/diversity-high-tech.

bias. Meanwhile, my results find limited support for the "biased programmer" hypotheses given I observe no differences by the gender, technical background, or educational institution of the manager who wrote the algorithm. Overall, these results suggest that mitigating algorithmic bias will require further work on understanding which modeling approaches can be effective in reducing adverse impacts in algorithms.

2.3 The returns to algorithmic decision-making

Finally, this paper is related to the large literature on the business impacts of algorithmic adoption. The information technology (IT) literature has paid special attention to the organizational practices and strategies that increase or decrease the returns to algorithmic decision making and IT investments more broadly. Since the 1990's, this literature has examined the set of practices that are complementary to IT adoption (see Melville et al. (2004) for a review). These include decentralized decision authority (Hitt and Brynjolfsson 1997; Bresnahan et al. 2002; Wu et al. 2019), changes in business strategy (Bartel et al. 2005), improved screening in hiring (Hitt and Brynjolfsson 1997; Bresnahan et al. 2002), employee training (Bapna et al. (2013)), the degree of external focus (Tambe et al. 2012), and performance pay (Aral et al. 2012; Dixon et al. 2021), amongst others.

One particular focus has been the human resource management practices that are complementary to algorithmic decision making and analytics. Tambe (2014) shows that the returns to big data technologies require significant data assets and labor markets where many workers have the required big data technology skills. Using U.S. Census Bureau data on manufacturing establishments, Brynjolfsson et al. (2021) illustrates that the returns to prediction technology are larger for organizations with more educated workers and with better managerial capacity. Rock (2021) shows that having hired workers with AI skills increased the returns to deep-learning technology by using Google's open-source launch of TensorFlow as a natural experiment. Overall, these papers have illustrated that the returns to algorithmic decision making and analytics depend critically on the human resources that a firm has.

My paper contributes to this literature by examining whether and how various hiring strategies are complementary to algorithmic decision-making. My set-up allows me to generate an estimated treatment effect of algorithmic decision making for each manager in my sample. I can then relate this treatment effect to a few human resource characteristics, including whether the manager has a technical

degree, machine learning skills[27], or is from an elite university. In my results, I show that hiring for technical talent and hiring for machine learning skills are complementary to algorithmic decision making. Technical workers with machine learning skills vastly outperform technical workers without these skills due to the increase in the predictive accuracy of their algorithms. Meanwhile technical workers with machine learning skills do better than non-technical workers with machine learning skills, likely by being more careful with implementation. Meanwhile, selective hiring and technical hiring have no individual impact on the returns to algorithmic adoption, nor a complementary one. Instead, firms that hire technical workers with machine learning skills see the largest gains from algorithmic adoption. These results point to an important complementarity between hiring for technical workers and for machine learning skills, and suggest that organizational redesign in the age of algorithms will require firms simultaneously screen for both. More broadly, it suggests that organizations that adopt data driven decision making change their job demand characteristics toward workers with more technical skills and more technical backgrounds.

3 Setting

The task occurred as a mandatory assignment in a People Analytics course at a top MBA program. The course was held eight times from 2019 to 2022.

3.1 Task

The task involved managers designing their own hiring algorithm to improve hiring at a firm. The managers received data on around 2,000 workers, 689 of whom were currently employed at the company, and 1,258 of whom were who were not currently employed but were eligible to be hired. Managers had access to worker's demographic variables (such as gender, age, and education), employment characteristics (such as work status), and current employer characteristics (such as firm size and workplace flexibility). For the 689 hired workers, the managers had a measure of job performance that was unbiased and without noise. The task required managers to analyzing this data to generate recommendations for the company to improve its hiring.

[27]This mirrors the differential impacts of predictive (Brynjolfsson et al. 2021) versus descriptive analytics (Berman and Israeli 2021). In my context, I show that the value to predictive technologies versus descriptive ones is much higher, at least in the context of hiring.

The task consisted of two parts. In the first part, the managers had to examine the workforce data and make recommendations on how the company could improve its hiring practices. The managers first had to examine the predictors of being hired at firm F (P1Q1) and of job performance (P1Q2). Participants were not required to use a specific type of analysis, but most used linear and logistic regression (see section 5.1 for analysis of the methods used). Question three then asked participants to formulate a hypotheses, using their previous analysis, regarding how the firm could improve its hiring practices (P1Q3). For example, if a manager found that men were more likely to be hired but there were no gender differences in job performance, they could argue that the firm could improve its hiring practices by placing no weight on gender in allocating interviews. Similarly, they could use propose using the predicted values from the job performance regression to select candidates to interview. In this question, managers generated a set of rules to be followed in determining which candidates to hire. In the final question of part one, the managers had to use this rule to come up with two lists of 20 workers (P1Q4). First, they had to generate a list of 20 workers who were hired by the firm, but would have been rejected if the firm was using the hiring algorithm from P1Q3. Second, they had to generate a list of 20 applicants from the applicant pool who were not initially hired, but would have been hired had the firm used the hiring algorithm from P1Q3.

In the second half of the assignment, participants had to evaluate the performance of their algorithms and make a recommendation for or against adoption. Following the submission of all parts of part 1, the managers received performance data for all applicants (not just those initially hired by the firm). They then tested whether the twenty applicants they proposed selecting performed better or worse than the twenty workers they proposed rejecting (P2Q1). They then examined the impact of their proposal on workplace diversity (P2Q2). For example, they could analyze whether their hiring algorithm increased or decreased gender diversity at the firm. The prompt was intentionally kept broad so that the managers could examine diversity across their desired dimension. Finally, the manager was asked whether they would recommend that the firm adopt this algorithm, assuming there were no alternatives besides their algorithm and continuing with the status quo (P2Q4). The manager then had up to 300 words to describe their reasoning for their decision. Appendix B displays a copy of the instructions that participants received. The data that managers received came from the PIAC dataset (Cowgill et al. 2020b), an international survey that measures cognitive and workplace skills in over 40 countries.

3.2 Benefits of the setting

There are a number of reasons why this task presents a natural setting to study how the demographic representation impacts of algorithms shape their adoption decisions. First, the setting allows me to hold constant other factors that may impact algorithmic adoption. In the previous paragraphs, I documented that various factors influence the rate of algorithmic adoption, such as the amount of training data available or the existence of a measurable job performance outcome. My setting holds the input data factors that increase the benefits of algorithmic adoption, while providing a task where the statistical challenges are already dealt with. This allows me to zero in on how demographic representation shapes algorithmic adoption recommendations, which would be difficult in settings that feature variation in data inputs or differences in statistical measures.

Second, the setting provides a way to study and measure differences in algorithmic adoption. It is notoriously difficult to get data on firm technology adoption decisions, and relate them to performance outcomes. Such studies have either use protected Census data (see, for example, Brynjolfsson and McElheran 2016a,b; Brynjolfsson et al. 2021), large sample surveys initiated by researchers (see, for example, Hitt and Brynjolfsson 1997; Bresnahan et al. 2002; Brynjolfsson et al. 2003), or case studies in narrowly-defined industries (see, for example, Ichniowski et al. 1997; Bartel et al. 2007; Berman and Israeli 2021), which are difficult to access or may struggle with external validity issues. My setting contains a clean measure of algorithmic adoption recommendations, which I can then relate to the performance outcomes of algorithms. Although these are adoption recommendations and not decisions, they are still important inputs into algorithmic adoption, and can shed light on how demographic considerations shape the use of algorithms.

Third and finally, this setting offers a way to bypass the Fundamental Problem of Causal Inference and estimate the causal impact of algorithms on firm outcomes. Understanding how the effects of an algorithm impact adoption rates requires an unbiased measure of how the algorithm changes job performance and demographic representation. While it is possible to estimate how a given algorithm would influence demographic diversity (for example, by using algorithmic audits as described earlier), such audits are more difficult for job performance since this measure is only collected conditional on having been hired. Doing so for job performance runs into the Fundamental Problem of Causal Inference (Rubin 1974; Holland 1986), which states that researchers can never observe unit-level causal effects because we can only observe one potential outcome per unit. For example, if we wanted to

make a statement about the causal impact of adopting a hiring algorithm on job performance, we only observe one potential outcome: we observe the treated potential outcome if the firm adopts the algorithm, and we observe the control potential outcome if the firm does not adopt. The task allows me to avoid this problem by simulating each managers' potential outcomes when algorithmic hiring is implemented versus when it is not. More specifically, each manager's potential outcome under the status quo (without algorithmic hiring) is the average outcome Y across the 689 employees who are employed at the firm at the start of the assignment. Meanwhile, the managers' potential outcome under algorithmic hiring is the average Y across the 689 workers, after removing the 20 workers who the manager's algorithm determined should have never been hired, and adding the 20 workers who the manager's algorithm determined should have been hired. I can then compare outcomes across these two pools to measure the causal impact of algorithmic hiring for each manager in my sample, for various outcome measures including job performance and worker representation. Although no proposal was ever implemented, I can estimate a unit-level causal effect of adopting a hiring algorithm on firm outcomes.

For these reasons, the task presents a creative setting that bypasses many of the difficulties in studying algorithmic adoption decisions.

3.3 Participants

Overall, 397 participants completed the assignment across eight sections from 2019 to 2022. Appendix A.1 displays the number of participants who completed the assignment by section. In Table 1, I examine summary statistics of this population. 41% were female. In terms of education, 87% completed their undergraduate degree in the US, with 14% earning a BS. Ten percent went to a top–25 undergraduate institution. Around a third were in the EMBA program. The industry distribution was 26% in finance, 19% in technology, 17% in business services, 12% in social services, 9% in arts, 5% in real estate, and 12% in other industries. These numbers line up with the backgrounds of managers in the United States.

3.4 Data

My data consists of a few sources. First, I have data on the background of participants. This was collected directly from resumes that were submitted prior to the first week of class. This includes their educational background and work experience. Second, I have data on each participants' assignment submission. This includes the method and analysis conducted for P1Q1 and P1Q2, the hiring recommendation from P1Q3, the proposed candidates to hire/not hire from P1Q4, and the decisions to incorporate their proposal or not from P2Q3. Finally, I have the raw data that participants received to conduct the assignment. This allows me to merge manager hiring recommendations with candidate performance and demographic information to study the exact nature of tradeoffs that the managers were facing.

4 Empirical strategy

The goal of this paper is to examine the impact of various hiring algorithms on the kinds of candidates that are hired, and then examine how these impact whether the algorithm are adopted. To that end, the specifications in this paper come in one of a few forms.

First, I want to understand how each manager's hiring algorithm impacted the job performance and the demographic makeup of new hires. In order to do this, I run the following regression:

$$y_{m,c} = \beta_0 + \beta_1 * Accepted_{m,c} + \delta * X_m + \epsilon_{s,c} \tag{1}$$

where m indexes managers and c indexes job candidates. $Accepted_{s,c}$ is a binary indicator that equals one if manager m's algorithm accepted job candidate c, and is zero if it rejected this candidate. $y_{s,c}$ represents various outcome measures, including job performance, gender identity, age, education and region of origin. X_m is a matrix of manager-level controls including their gender, education, and prior industry, and programming method used. In some of the results, I estimate equation 1 but replace X_m with M_m, a set of manager-level fixed effects. This looks for variation within each manager, comparing the twenty candidates their algorithm accepted versus the twenty candidates their algorithm said to rejected. I estimate equation 1 with robust standard errors clustered at the manager

level. The coefficient of interest is β_1, which measures the average difference in outcomes for candidates hired versus rejected by the algorithm.

In my second analysis, I model how the impacts of the hiring algorithm affect whether or not it is adopted by the manager. To do so, I estimate various versions of the following model:

$$Adopt_m = \alpha_o + \alpha_1 * AdverseImpact_m^{gender} + \theta * A_m + \epsilon_m \tag{2}$$

where m indicates managers. $Adopt_m$is a binary indicator that equals one if manager m suggested adopting their hiring algorithm, and zero otherwise. $AdverseImpact_m^{gender}$ is a binary indicator that captures whether the hiring algorithm considered by manager m decreased female representation at the firm. It equals one if manager m's hiring algorithm reduces the number of female hires, and zero otherwise. A_m consists of various controls at the manager/algorithm level, including $AdverseImpact_m^{age}$ and $AdverseImpact_m^{region}$ (defined similarly to $AdverseImpact_m^{gender}$) and $PerformanceImpact_m$, the percent change in job performance for hires from the algorithm written by manager m. I estimate equation 2 with robust standard errors. The coefficients of interest is α_1, which measure the impact of an adverse gender impacts on the likelihood of recommending the adoption of the hiring algorithm.

Finally, I examine how algorithmic inputs and manager identity predict whether an algorithm decreases female representation. To do so, I estimate the following model:

$$AdverseImpact_m^{gender} = \lambda_0 + \lambda M_m + \epsilon_m \tag{3}$$

M_m is a vector of variables at both the manager and hiring algorithm level. The manager characteristics include gender, education, and prior industry. The algorithmic characteristics include the type of modeling strategy used and the model's data inputs. The coefficient λ will capture whether there are certain manager or algorithmic features that predict finding solutions that have no negative impact on the number of female hires.

5 Empirical results

5.1 Overview of modeling strategies

I begin by examining the content of the hiring algorithms that were created. The bottom of Table 1 examines the distribution of modeling strategies across the whole sample. The most popular approach, used by over 80 percent of managers, was linear regression. These managers regressed job performance on various candidate covariates, and then used the predicted values from this regression to generate predicted performance values for all job candidates. The second most common approach was tabulation, which was used by 13 percent of managers. Managers who used this approach examined average job performance across different values of the demographic variables. For example, a manager would look at average job performance across different education categories and form a hypothesis to accept candidates from the bucket with the highest average job performance. A few students (four percent) used machine learning techniques such as a random forest or having a training and testing set. The managers also differed in their use of demographic variables, and their predictions. Fewer than 30 percent were mindful of using sensitive demographic data such as gender, age, and geographic region. Managers also differed in the number of empirical models they considered. Around half of the managers considered multiple models in their approach (for example, running multiple regression models and choosing the one with the highest R^2). Overall, the subject population used relatively strong empirical techniques in the design of their hiring algorithms.

5.2 The impact of the hiring algorithm on performance and demographics

I next examine the impact of the proposed hiring algorithm on the firm's hires. Recall that each manager's algorithm generated a sample of 20 candidates who were rejected by the firm but should have been hired if the algorithm was used, and a sample of 20 candidates who were hired by the firm but should have been rejected according to the algorithm. The impact of adopting the algorithm can thus be examined by comparing average outcomes for the 20 accepted candidates versus the 20 rejected candidates.

5.2.1 Impact on job performance

I first examine the impact of the hiring algorithm on the job performance of accepted candidates. Table 2 displays the results of equation 1 using job performance as the outcome measure. Column 1 compares accepted versus rejected candidates, while column 2 adds manager controls such as the gender, education, and prior industry of the manager. Meanwhile, columns 3 replaces the manager controls with manager fixed effects.

My results indicate that the hiring algorithms on average led to large increases in the job performance of accepted candidates. The twenty accepted candidates had a job performance score that was over 470 points higher on average than rejected candidates. This difference is large and economically significant: The average job performance of the rejected candidates was 2,328, indicating that the algorithm led to a 20 percent improvement in job performance relative to the control group.[28] This is also true not just for the average candidate but also throughout the performance distribution of recommended candidates. In Appendix A.1.1, I run quantile regressions to examine the impact of the hiring algorithm throughout the distribution. Results from quantile regressions reveal that the algorithms increases the performance of the bottom 10 percentile of hires by over 30 percent, and the top 10 percentile of hires by around 13 percent.

In addition to examining the average impact of the hiring algorithm across the distribution of managers, I can also examine manager-specific treatment effects. This allows me to understand the extent to which all managers were able to improve performance, or whether there was large heterogeneity in treatment effects. I do so by estimating equation 1 for each manager, and then converting the estimated treatment effect into a percent improvement (by dividing the estimated coefficient by the average performance of rejected candidates for each manager). I plot these treatment effects and confidence intervals in Figure 1, sorting the effects from largest to smallest. The figure illustrates that most were able to improve performance outcomes. Over 80 percent of managers wrote a hiring algorithm that increased the job performance of hires; among these, the average increase was 32 percent. Those whose algorithms decreased job performance did so by, on average, 11 percent. Overall, the managers were highly effective in writing hiring algorithms that improved the job performance of hires.

[28] If I compare the size of this treatment effect to the entire pool of hired workers, it represents an effect of 17 percent (= 478.1/2832.6).

5.2.2 Impact on demographic representation

I next examine the impact of the hiring algorithm on the demographic distribution of hires. I do so for three protected categories. First, Table 3 displays the results of equation 1 using a binary indicator for being a female candidate as the outcome measure. The coefficient captures the percentage point change in the number of female hires: a positive number signifies that the algorithm will lead to more female hires than the status quo. The results in Table 2, however, show that the hiring algorithms on average have no impact on the gender representation of hires. In fact, the point estimate is negative, signifying that the hiring algorithms, if anything, reduce the number of female hires, though this effect is not statistically significant at conventional levels. Although the algorithms contain no impact on the gender of hires, this conceals vast heterogeneity by manager. Figure 2 examines the impact by manager. This is similar to Figure 1 except the x-axis displays the impact on the raw number of female hires rather than the percent change. Slightly over half of the hiring algorithms increase the number of female hires, while the other half decrease it.

The hiring algorithm also led to changes in the distribution of hires relative to two other protected categories: age and region of origin. Table 4 displays the results of equation 1 for various age (Panel A), and region (Panel B) brackets. The algorithm shifted the age distribution of the firm downward. Hired workers were 40 percent less likely to be aged 50–65, while the proportion of workers between the ages of 16 and 24 doubled. In terms of education, there was three-fold increase in the number of hires with above a high school level of education. These came at the cost of fewer hires with a high school or below a high school education. The hiring algorithms also shifted the regional allocation of hires. There were large increases in the number of candidates from Asia, North America, and Western Europe, and fewer candidates from Latin America.

5.3 What predicts the adoption of hiring algorithms?

In this section, I examine how the effects documented in sections 5.2.1 and 5.2.2 impact a managers' adoption recommendation. Table 5 displays the results of equation 2 using various control variables in A_m. Column 1 examines the relationship between lowering female representation and the decision to adopt. Column 2 adds a control variable for the impact of the hiring algorithm on job performance, while column 3 additionally includes binary indicators for having an adverse impact on age and country

of origin, respectively. Finally, column 4 adds manager-level controls, including gender, student type, educational background, and work experience.

The results in Table 5 indicate that hiring algorithms that decrease female representation at the firm are less likely to be adopted. Column 1 indicates that having an adverse impact on female representation lowers the rate of adoption by 35 percentage points. Since 74 percent of hiring algorithms with no adverse impact are adopted, this effect translates into a 47 percent decreases in the likelihood of adoption. This effect remains unchanged if I control for the algorithm's impact on job performance (column 2), and slightly increases in magnitude when controlling for adverse impacts on other demographic characteristics (column 3) and manager controls (column 4). The effects documented in Table 5 are also economically meaningful. Comparing the coefficient on the female representation indicator with the one on the performance measure allows me to examine how managers trade off increases in performance versus decreases in female representation. In order for a hiring algorithm that decreases the number of female hires to be adopted at the same rate as one that does not, it would need to additionally increase the performance of hires on average by 94 percent more ($= \frac{0.346}{0.00369}$). Less than two percent of managers write algorithms with at least this large of a performance impact (see Figure 1), suggesting that adverse impacts on the number of female hires play a significant role in lowering algorithmic adoption recommendations.

5.4 Mechanisms

In the previous section, I documented that managers are unwilling to adopt hiring algorithms that decrease the number of female hires, and are willing to sacrifice large productivity gains to do so. In this subsection, I consider various mechanisms behind this effect.

5.5 Female representation loss aversion

I propose that "female representation loss aversion" can explain managerial adoption recommendations in my setting. Under this theory, managers have an aversion to implementing *any* hiring algorithm that leads to fewer female hires than the status quo, regardless of the size of the effect. A big literature in behavioral economics examines loss aversion, or the cognitive bias whereby humans prefer avoiding losses to acquiring equivalent gains ("losses loom larger than gains") (Kahneman and Tversky

1979; Brown et al. 2022). In the context of hiring algorithms, loss aversion with respect to female representation occurs if managers are averse to implementing hiring algorithms that are legal but lead to even a small decrease in the number of female hires. Moreover, this loss aversion is asymmetric because the effect is concentrated amongst hiring algorithms that decrease the number of female hires, and not male hires.

In order to test the "female representation loss aversion" hypothesis, I use a regression discontinuity design (Angrist and Lavy 1999; Hahn et al. 2001; Lee and Lemieux 2010). If female representation loss aversion was driving managerial adoption recommendations, then I would expect that managers are less likely to adopt hiring proposals that marginally decrease the number of female hires, versus algorithms that do not change the gender distribution of the firm or marginally increase the number of female hires. I test this hypothesis by comparing adoption recommendations for hiring algorithms that decrease the number of female hires by one or two, versus algorithms that have no impact or increase the number of female hires by one or two. In Appendix A.2, I show that hiring algorithms to the right versus left side of this cutoff exhibit no systematic differences in their job performance or demographic impacts on age or country of origin. This suggests that the regression discontinuity design will provide a valid estimate of the causal impact of marginally reducing female representation on algorithmic adoption recommendations.

Table 6 examines differences in algorithmic adoption recommendations along this cut-off. Although my sample size here is limited, the results indicate that hiring algorithms that narrowly decrease female representation were 33 percent less likely to be adapted than those that had no impact or marginally increased it. Moreover, this effect size is large. Table 5 shows that algorithms that decreased female representation were 45 percent less likely to be adapted. This means that around three-quarters of the overall effect of lowering the number of female hires comes from the presence of loss aversion ($= 33.3/45.2$).

In Appendix section A.2, I conduct a more specific test of the "female representation loss aversion" hypothesis. That section compares adoption recommendations for algorithms that marginally increase or marginally decrease female representation relative to the control group of no change. If the "female representation loss aversion" hypothesis was true, we would see that marginal decreases in female representation were more costly in terms of adoption recommendations than the benefits to algorithms with a marginal increase in female representation. The results in Appendix section A.2 indicate that

algorithms that weakly increase female representation were just as likely to be adopted as those that had no impact, while those that marginally decrease female representation were much less likely to be adopted relative to both types, providing evidence that managers showed asymmetric responses to algorithms that decreased female representation in their recommendations.

Overall, these results indicate that managerial adoption recommendations are strongly shaped by an aversion to even marginal decreases in female representation.

5.6 Alternate explanations

In this section, I rule out various alternate explanations.

5.6.1 Avoiding algorithms that discriminate against protected class

The first alternate explanation is that managers do not adopt these algorithms because they fear legal repercussions for discriminating on the basis of a protected class. According to this theory, managers are weary of adopting algorithms that lead to adverse impacts on protected categories because they fear the repercussions of legal action.

However, a few patterns in the data hint that legal fears are not the primary driver of adoption decisions. First, if legal concerns regarding discrimination were the primary reason why these algorithms are not adopted, then managers would be unlikely to adopt hiring algorithms that discriminate against any protected class. However, as the results in Table 4 show, this is not the case. Although age and country of origin are protected classes, adverse impacts in these domains do not influence manager adoption decisions. In fact, the results seem to tolerate some level of adverse impact in region of origin for adoption decisions.

There is a second and more direct test that can rule out legal concerns as the primary motivation. The U.S. Equal Employment Opportunity Commission (EEOC) has a rule of thumb to determine whether adverse impact exists within a given selection device: the four-fifths rule.[29] This rule states that a selection rate for any group that is less than four-fifths of that for the group with the highest rate constitutes evidence of adverse impact (also called 'disparate impact'), that is, discriminatory

[29] See https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines for more information.

effects on a protected group. If managers were concerned of legality, they would be particularly wary of adopting hiring algorithms that violate the 4/5ths rule versus those that do not violate the 4/5ths rule (but still decrease the number of female hires).

In order to investigate this possibility, I conduct an adverse impact analysis for each manager's algorithm. I calculate an average selection rate for men and for women using the 20 candidates they recommended, and the 1,258 candidates in the applicant pool (704 women and 554 men). The selection rate for female candidates is given by $\frac{N_F^{Selected}}{N_F^{Available}} = \frac{N_F^{Selected}}{704}$, where $N_F^{Selected}$ measures the number of female candidates selected by the manager, while the selection rate for men is given by $\frac{N_M^{Selected}}{N_M^{Available}} = \frac{N_M^{Selected}}{554}$. Adverse impact exists if the ratio of the female-to-male selection ratios is under 0.80, or if the male-to-female selection ratio is under 0.80. Appendix A.1.3 displays the a histogram of these adverse impact coefficients. 78 percent of algorithms failed the $\frac{4}{5}^{th}$'s rule, and 60% of these had an adverse impact for women. However, managers still suggested adoption in over 70 percent of the algorithms that featured an adverse impact for men (see Table A.3 in Appendix Section A.1.3). Moreover, at least manager did not suggest adoption for a hiring algorithm that decreased the proportion of female hires but did not violate the $\frac{4}{5}^{th}$'s rule. For these reasons, it is unlikely that avoiding illegal discrimination was the primary driver of managerial decision-making.

In Appendix sectionsubsec:'s-rule-analysis I further rule out that adoption decisions are driven by avoiding discriminatory algorithms.. Appendix Figure A.4 displays the average recommendation rates for algorithms based on whether they (i) decrease female representation or not, and (ii) would pass or fail the four-fifths test. If fear of illegal behavior were driving my results, I would expect to see that algorithms with an impact ratio of less than 0.80 are less likely to be adapted than those with an impact ratio of 0.80 or more. However, this is not the case. The results indicate that passing or failing the four-fifths test has no bearing on the likelihood of adoption Instead, my results indicate that decision-makers adopt algorithms that increase female representation, regardless of whether they would be legal or not, and avoid algorithms that decrease female representation, regardless of whether they would be legal or not. For these reasons, my results are unlikely to be driven by legal fears.

5.6.2 Incentives

A second alternate explanation is that the managers in my sample were incentivized, either explicitly or through social pressures, to consider the demographic impact of an algorithm in their adoption

recommendations. There are a variety of ways that this could occur. For example, if the managers were primed to arguments or discussions of algorithmic bias and its inescapability, they may be more likely to avoid using the algorithm.[30] This could also occur through the grading incentives managers faced; those in my sample may be less likely to suggest algorithmic adoption if they believed doing so would risk a high-quality grade. Finally, there may be social pressures to respond in support of considering the demographic impacts of algorithms (for example, if fellow classmates would see their responses). Overall, these concerns would suggest that the results in this paper are due to unique features of the task and setting, rather than a more general decision-making process.

There are various aspects of the task and setting that mitigate these concerns. First, the course included a balanced discussion of the merits and drawbacks of algorithms and bias, and a case study where algorithmic hiring improved diversity at an organization.[31] Second, the instructions made clear that the evaluations would be based on the quality of the argument for or against adoption, and not the actual recommendation decision. Third, managers were not told of the specific recommendations of others. Instead, results from the assignment were only shared in aggregate (for example, X percent of the class suggested adoption). For these reasons, it is unlikely that incentivization through unique features of the setting drove the impacts I observe.

6 Discussion

6.1 What predicts whether a given algorithm will decrease female representation?

The results in the previous sections documented that hiring algorithms that decrease female representation are much less likely to be adopted. In this subsection, I examine what algorithmic and managerial features predict whether a given algorithm will decrease female representation or not. As described in Section 1, there are a few approaches to consider; here I focus particularly on two sets of predictors: the algorithmic approach used, and the identity and background of a given coder. In terms of the algorithmic approach used, I examine the modeling strategy (uses machine learning versus regression

[30] Indeed, Cowgill et al. (2020a) provide evidence that decision-makers are less likely to suggest algorithmic use when given arguments regarding the inescapability of algorithmic bias.

[31] If anything, it seems like the bias here would go in the opposite direction. The course emphasized that while concerns about algorithmic bias are legitimate, human decision makers are oftentimes more biased. Meanwhile, algorithms enable auditability, which is much more difficult for human decision makers. This would lead decision-makers to favor algorithms over human decision-making if they were concerned about the distributional impacts of selection processes.

versus tabulations), the input data (whether the algorithm used sensitive covariates or not), and testing approach (whether the algorithm featured multiple models or not). These will determine whether there are any algorithmic techniques that decrease the likelihood of an adverse gender impact regardless of the manager who wrote the algorithm. Meanwhile, the manager attributes I examine include their gender, education, and experience. These capture whether coder demographics have an impact on the likelihood of decreasing gender representation above and beyond their effect on the algorithmic approach used. Table 7 displays these results by regressing an indicator for whether or not a given algorithm decreased gender representation on various covariates. Column 1 examines what programming techniques influence the likelihood of decreasing the number of female hires, while Column 2 examines how managerial covariates impact this likelihood. In column 3, I examine them both in the same model.

These results indicate that certain algorithmic methods are useful in predicting whether or not a given algorithm will decrease the number of female hires. One common hypothesis is that blocking the use of sensitive information such as gender in algorithms can ameliorate adverse gender impacts. Table 7 finds some support for this. Algorithms that do not use sensitive demographic variables in their predictions are 13 percentage points less likely to decrease female representation ($p = 0.03$). Using multiple models also seems to lower the likelihood of decreasing female representation, though the magnitude of this effect is about half as small and not statistically significant at conventional levels. Meanwhile, I find some evidence that using regression increases the likelihood of finding an adverse gender impact relative to tabulations.[32] These results lend credibility to the argument that certain modeling strategies can be effective in reducing the likelihood of decreasing gender representation in an algorithm .

Meanwhile, the demographic characteristics of the manager who wrote the algorithm had a more muted effect on the likelihood of decreasing female representation. Algorithms written by female managers were slightly less likely to decrease the number of female hires, though this effect is not statistically significant ($p = 0.41$).[33] Technical expertise or academic background did not seem to matter much: algorithms written by those with BS degrees or from top universities were as likely to find an adverse female representation as those without a BS degree and those from non-elite universities, respectively. Overall, these results suggest that modeling approaches may be useful in mitigating

[32]This is likely because tabulations can more easily avoid using gender in their predictions.
[33]This lines up with the results in Cowgill et al. (2020c).

the likelihood of decreasing female representation through the use of algorithms., while programmer characteristics are less

6.2 Complementarities between human resource management practices and algorithmic decision-making

Another interesting question is whether there are human resource strategies that are complementary to algorithmic adoption. Although the average impact of algorithms is often positive, firm practices likely moderate the average return to algorithmic decision-making. Indeed, while Table 2 shows that the average impact of the algorithm on the job performance of hires is positive, Figure 1 illustrates that this estimate conceals a lot of heterogeneity. Understanding the source of this heterogeneity can help determine how algorithmic decision-making is related to a firm's broader strategy.

In this subsection, I examine whether the use of algorithms in decision-making is complementary to three different human resource strategies: (i) selective hiring (whereby the firm only seeks to hire candidates from top universities); (ii) technical hiring (whereby the firm only seeks to hire candidates who have an undergraduate degree in a B.S. field); and (iii) machine learning hiring (whereby the firm only seeks to hire candidates who have machine learning skills).

My goal is to relate these features to the returns of algorithmic decision-making. An ideal study would randomize algorithmic decision-making across firms that also vary in the extent to which they rely on selective and/or technical and/or machine learning hiring (possibly also through random assignment). However, this is not possible. Using data from managers in my sample, meanwhile, is complicated by two factors. First, no hiring algorithms were ever adopted in my setting given that this was a class assignment. Second, my data is at the manager, and not the firm level, which makes it difficult to make strong statements regarding the impact of various hiring strategies on the returns to algorithmic decision-making.

I assuage these concerns in the following way. First, it is possible to use the data at hand to estimate the causal impact of algorithmic adoption if it were adopted at Firm F. Recall that it is possible to estimate a manager-level treatment effect of algorithmic adoption on the job performance. In the introduction and empirical strategy, I discuss how I can simulate each manager's treated potential outcome (average outcomes at the firm if they implement algorithmic hiring) and their control potential

outcome (average outcomes at the firm if they maintain the status quo). The impact of algorithmic hiring for manager m is simply the difference between these two quantities, or:

$$\tau_m = E[Y_m|UsesAlgorithms] - E[Y_m|KeepsStatusQuo].$$

For each manager in my sample, I thus have a measure of τ_m, or the causal impact of adopting the hiring algorithm written by manager m.

Second, in order to make inferences about complementarities between firm hiring policies and the impact of algorithmic hiring τ_m, I assume that each manager m corresponds to a hypothetical firm with the set of hiring practices that lead the firm to hire manager m. I index firms by (S_m, T_m, M_m), where $S_m = 1$ if the firm has a selective hiring policy (so that manager m went to an elite university if $S_m = 1$) and 0 if they do not (when manager m did not go to an elite university). Similarly, $T_m = 1$ if the firm hires technical talent (so that manager m has a B.S. degree if $T_m = 1$) and 0 if they do not (when manager m does not have a BS), with a similar definition using $M_m = 1$ if the firm screens for machine learning skills, and 0 otherwise. I can then estimate the relationship between τ_m and (S_m, T_m, M_m) to understand whether there are complementarities between firm screening practices and the returns to algorithmic decision-making.

While this setup may feel unrealistic, it does mimic how complementarities are studied in the field. A common approach is to restrict sample observations to very specific types of tasks in order to eliminate sources of heterogeneity that could confound the effects studied. For example, in their study of complementary human resource management practices, Ichniowski et al. (1997) examines output in a set of 36 homogenous steel production lines. Similarly, Bartel et al. (2007) limit their attention to valve manufacturing plants to examine how information technology influences productivity. In a similar light, I examine how different human resource management practices impact the returns to algorithmic decision-making in a narrowly defined firm, Firm F. Moreover, each manager's initial stock of employees and eligible candidates is the same, and they have access to the same prediction covariates. The only variation comes from the identity of the manager, which influences how the algorithm is written and thereby impact of algorithmic adoption on Firm's F performance. This thus presents a creative way to understand the complementarity between different firm screening practices and the returns to algorithmic decision-making.

Table 8 studies these complementarities by relating τ_m to various functional forms of (S_m, T_m, M_m).

The results indicate that hiring for machine learning skills has a large impact on the performance of hiring algorithms. Algorithms written using machine learning outperform those using regression and tabulations by around six percentage points. Meanwhile, selective hiring and technical hiring have no impact on the returns to algorithmic adoption, and there is also no complementarity between selective hiring and the returns to technical hiring. Instead, I do find some evidence of a complementarity between technical hiring and hiring for machine learning skills (columns 4 and 5).These estimates show that technical workers with ML skills outperform both technical workers without ML skills, and non-technical workers with machine learning skills. These results illustrate that hiring for technical workers and for machine learning skills is complementary to algorithmic adoption.

7 Conclusion

This paper is motivated by the rise of algorithms in hiring. Various business and labor market trends have led to a secular rise in the demand for algorithmic hiring. These include improved predictions driven by advances in machine learning, rising application volumes due to job search platforms, and increasing returns to selective hiring. Each of these trends has increased the relative attractiveness of using algorithms to algorithms to sort and select the workers

This rapid increased prevalence of algorithms, however, has made questions concerning the impact on candidates and the firm more pressing. Hiring algorithms will likely shift not just the job performance of a firm's hired workers, but also their demographic characteristics. This paper examines how tradeoffs between performance and various diversity considerations impact algorithmic adoption recommendations. Using data from the hiring algorithms and proposed adoption recommendations from over 450 business managers and executives, I find that gender diversity considerations play an important role in shaping algorithmic adoption decisions. Algorithms that decreased the number of female hires were half as likely to be adopted as those that had no impact or increased it, while adverse impacts on age and region of origin, two other protected classes, had no effect on adoption decisions. I identify "female representation loss aversion" as an impediment to algorithmic adoption, whereby managers are less likely to suggest adopting hiring algorithms that marginally decrease the number of female hires versus lead to no change or weakly increase them. Overall, the paper highlights the unique role played by gender diversity in shaping algorithmic adoption decisions.

This paper suggests three avenues for future research. First, this paper opens up the question of how algorithmic hiring influences the candidate pool available to organizations. While my paper examines how algorithmic hiring impacts the types of candidates that are selected (conditional on a given applicant pool), an equally interesting question is how algorithmic hiring shifts the type of workers in a firm's applicant pool. There is a growing literature that examines how firm attributes impact labor supply (see, for example, Burbano 2016 and Abraham and Burbano 2022). If workers draw a disunity from being screened algorithmically, using algorithms to hire will shift the set of workers from whom to hire. Indeed, a recent Pew survey indicates that only 22 percent of U.S. adults say they would be comfortable applying for a job where an algorithm would make hiring decisions.[34] Men and younger workers were more likely to support algorithmic hiring, suggesting that algorithmic hiring may skew a firm's applicant pool to be predominantly young and male. Moreover, to the extent that workers gain a non-pecuniary benefit (Cassar and Meier 2018) from being screened by a human, firms would need to design incentives to increase applications to jobs with algorithmic screening. An exciting direction for future research is to study whether algorithmic hiring influences who applies, and the willingness of candidates to provide labor at a given salary level. For example, if algorithmic hiring shifts the pool of applicants to lower-quality workers, or applicants need to be compensated for being screened by a machine, the returns to algorithms in hiring may be more limited such that adoption is disadvantageous.

Second, there is more work to be done on firm differences in the potential tradeoffs of algorithmic hiring. This paper examines how managers consider performance and diversity tradeoffs in their adoption recommendations, but it would also be interesting to examine how firms evaluate these trade-offs, and how algorithms fit with their broader strategies. These considerations are gaining more prominence for governments and the general public. Data on the hiring practices of firms is rare (Oyer and Schaefer 2010), but data concerning algorithm hiring may become available soon. For example, New York City just signed a law that states that firms that use algorithms in their hiring process are required to not only audit the results of their algorithms, but also post the results of their audits on their websites.[35] Such regulations present researchers with creative opportunities to get data on the use of algorithmic hiring in firms. Moreover, variation in how firms collect and report their data present an exciting opportunity to study company disclosures, both in the empirical tests used by companies and also in the results of the audits themselves. These disclosures likely say a lot about a

[34] https://www.pewresearch.org/internet/2017/10/04/americans-attitudes-toward-hiring-algorithms/
[35] https://news.bloomberglaw.com/daily-labor-report/new-york-city-ai-bias-law-charts-new-territory-for-employers

firm's human resources strategy.

Finally, algorithmic hiring is only one of the many changes to hiring that are a result of digitization and advancements in information technologies. For example, digitization and job platforms have also led to outbound recruiting, whereby firms seek out candidates directly rather than waiting for them to apply (Carrillo-Tudela et al. 2015; Black et al. 2022; Kim and Pergler 2022). Additionally, online labor markets have made it easier for firms to outsource and delegate hiring decisions to (human) third-party specialists (Cowgill and Perkowski 2022; Kolhepp and Aleksenko 2022). Firms thus have new strategic options when compiling their applicant pools (through inbound recruiting or outbound recruiting) and evaluating those pools (in-house by a human, in-house by an algorithm, outsourced to a human, or outsourced to a machine). An exciting avenue for future research is to relate a firm's broader hiring strategy choices to its internal structure, and characteristics about the job and pool of applicants.[36] For example, firms may be more likely to use algorithms in jobs with many applicants (where evaluating each applicant by a human is too costly) and more measurable performance metrics (where algorithms can do well at predicting who will be a good hire). Formalizing such hypotheses and testing them empirically can substantially increase our understanding of how firms screen and compete for talent in the digital age.

[36]There is also likely variation in the types of algorithms that are deployed (for example, using algorithms to scrape data from resumes, versus using facial analysis to collect data from online interviews).

References

Abraham, Mabel and Vanessa Burbano, "Congruence Between Leadership Gender and Organizational Claims Affects the Gender Composition of the Applicant Pool: Field Experimental Evidence," *Organization Science*, 2022, *33* (1), 393–413.

Agan, Amanda and Sonja Starr, "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 191–235.

Agan, Amanda Y, Bo Cowgill, and Laura K Gee, "Salary History and Employer Demand: Evidence from a Two-Sided Audit," Working Paper 29460, National Bureau of Economic Research November 2021.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Information Economics and Policy*, 2019, *47*, 1–6. The Economics of Artificial Intelligence and Machine Learning.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 1999, *114* (2), 533–575.

Aral, Sinan, Erik Brynjolfsson, and Lynn Wu, "Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology," *Management Science*, 2012, *58* (5), 913–931.

Assael, Yannis M., Brendan Shillingford, Shimon Whiteson, and Nando de Freitas, "LipNet: End-to-End Sentence-level Lipreading," 2016.

Athey, Susan, Raj Chetty, Guido W. Imbens, and Hyunseung Kang, "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," NBER Working Papers 26463, National Bureau of Economic Research, Inc November 2019.

Bapna, Ravi, Nishtha Langer, Amit Mehra, Ram Gopal, and Alok Gupta, "Human Capital Investments and Employee Performance: An Analysis of IT Services Industry," *Management Science*, 2013, *59* (3), 641–658.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw, "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills*," *The Quarterly Journal of Economics*, 11 2007, *122* (4), 1721–1758.

Bartel, Ann P, Casey Ichniowski, and Kathryn L Shaw, "How Does Information Technology Really Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement and Worker Skills," Working Paper 11773, National Bureau of Economic Research November 2005.

Berman, Ron and Ayelet Israeli, "The Value of Descriptive Analytics: Evidence from Online Retailers," *Working paper*, 2021.

Bhatia, Nandil and Stephan Meier, " Are Customers 10x More Important to Firms than Employees? Empirical Analysis of Imbalance in Emphasis Between Two Stakeholders," *Working paper*, 2022.

Black, Ines, Sharique Hasan, and Rembrand Koning, "Hunting for Talent: Firm-driven Labor Market Search in the United State," *Working paper, Available online at* https://www.hbs.edu/ris/Publication%20Files/Hunting_for_talent-3_91b04534-b543-485b-91bd-732c8d030efd.pdf, 2022.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer, "Meta-analysis of Empirical Estimates of Loss Aversion," *Forthcoming in The Journal of Economic Literature*, 2022.

Brynjolfsson, Erik and Kristina McElheran, "Data in Action: Data-Driven Decision Making in U.S. Manufacturing," Working Papers 16-06, Center for Economic Studies, U.S. Census Bureau January 2016.

_ and _ , "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review*, May 2016, *106* (5), 133–39.

_ , Lorin M. Hitt, and Kim Heekyung Hellen Kim, "Strength in numbers: How does data-driven decisionmaking affect firm performance?," *Working paper*, 2003.

_ , Wang Jin, and Kristina McElheran, "The power of prediction: predictive analytics, workplace complements, and business performance," *Business Economics*, October 2021, *56* (4), 217–239.

Burbano, Vanessa C., "Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces," *Organization Science*, 2016, *27* (4), 1010–1028.

Burton, Jason W., Mari-Klara Stein, and Tina Blegind Jensen, "A systematic review of algorithm aversion in augmented decision making," *Journal of Behavioral Decision Making*, 2020, *33* (2), 220–239.

Cappeli, Peter, " Making the most of on-line recruiting," *Harvard Business Review*, 2001, *79* (3), 139–166.

Carrillo-Tudela, Carlos, Bart Hobijn, Patryk Perkowski, and Ludo Visschers, "Majority of Hires Never Report Looking for a Job," *FRBSF Economic Letter, Available online at* https://www.frbsf.org/economic-research/publications/economic-letter/2015/march/labor-market-turnover-new-hire-recruitment/, 2015.

Cassar, Lea and Stephan Meier, "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, August 2018, *32* (3), 215–38.

Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann, "Task-Dependent Algorithm Aversion," *Journal of Marketing Research*, 2019, *56* (5), 809–825.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan, "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review*, May 2016, *106* (5), 124–27.

Choudhury, Prithwiraj, Evan Starr, and Rajshree Agarwal, "Machine learning and human capital complementarities: Experimental evidence on bias mitigation," *Strategic Management Journal*, 2020, *41* (8), 1381–1411.

Cowgill, Bo, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening," *Working paper; Available online at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3584916, 2020.

_ and Catherine E. Tucker, "Algorithmic fairness and economics," *Working paper, Available online at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3361280, 2020.

_ **and Patryk Perkowski**, "Delegation in hiring: Evidence from a Two-Sided Audit," *Working paper; Available online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3584919*, 2022.

_ **, Fabrizio Dell'Acqua, and Sandra Matz**, "The Managerial Effects of Algorithmic Fairness Activism," *AEA Papers and Proceedings*, May 2020, *110*, 85–90.

_ **,** _ **, Sam Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau**, "Replication Data (A) for 'Biased Programmers or Biased Data?': Individual Measures of Numeracy, Literacy and Problem Solving Skill – and Biographical Data – for a Representative Sample of 200K OECD Residents," 2020.

_ **,** _ **, Samuel Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau**, "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," in "Proceedings of the 21st ACM Conference on Economics and Computation" EC '20 Association for Computing Machinery New York, NY, USA 2020, p. 679–681.

Cullen, Zoe B, Will S Dobbie, and Mitchell Hoffman, "Increasing the Demand for Workers with a Criminal Record," Working Paper 29947, National Bureau of Economic Research April 2022.

De-Arteaga, Maria, Artur Dubrawski, and Alexandra Chouldechova, "Learning under selective labels in the presence of expert consistency," *CoRR*, 2018, *abs/1807.00905*.

Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski, "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination," *Working paper; Available online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3746564*, 2022.

Dietvorst, B. J., J. P. Simmons, and C. Massey, "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err," *Journal of Experimental Psychology: General*, 2015, *144* (1), 114–126.

Dixon, Jay, Bryan Hong, and Lynn Wu, "The Robot Revolution: Managerial and Employment Consequences for Firms," *Management Science*, 2021, *67* (9), 5586–5605.

Doleac, Jennifer L. and Benjamin Hansen, "The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden," *Journal of Labor Economics*, 2020, *38* (2), 321–374.

Glikson, Ella and Anita Williams Woolley, "Human Trust in Artificial Intelligence: Review of Empirical Research," *Academy of Management Annals*, 2020, *14* (2), 627–660.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, September 2000, *90* (4), 715–741.

Greiner, Ben, Philipp Grunwald, Thomas Lindner, Georg Lintner, and Martin Wiernsperger, "Incentives, Framing, and Trust in Algorithmic Advice: An Experimental Study," *Working paper*, 2022.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Hansen, Benjamin and Drew McNichols, "Information and the Persistence of the Gender Wage Gap: Early Evidence from California's Salary History Ban," Working Paper 27054, National Bureau of Economic Research April 2020.

Hassabis, Demis, " Artificial Intelligence: Chess match of the century," *Nature*, 2017, *544*, 413–414.

Hertz, Nicholas and Eva Wiese, " Good advice is beyond all price, but what if it comes from a machine?," *Journal of Experimental Psychology: Applied*, 2019, *25* (3), 386–395.

Hitt, Lorin M. and Erik Brynjolfsson, "Information Technology and Internal Firm Organization: An Exploratory Analysis," *Journal of Management Information Systems*, 1997, *14* (2), 81–101.

Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986, *81* (396), 945–960.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *The American Economic Review*, 1997, *87* (3), 291–313.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kim, Danny and Mike Pergler, "Startup Hiring through Firm-Driven Search: Evidence from VFA," *Working paper*, 2022.

Kirkpatrick, Keith, "Battling Algorithmic Bias: How Do We Ensure Algorithms Treat Us Fairly?," *Commun. ACM*, sep 2016, *59* (10), 16–17.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 237–293.

_ , Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan, "Algorithmic Fairness," *AEA Papers and Proceedings*, May 2018, *108*, 22–27.

Kolhepp, Jacob and Stepan Aleksenko, "Delegated Recruitment and Hiring Distortions," *Working paper, Available online at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3905019, 2022.

Kuncel, Nathan R., Deniz S. Ones, and David M. Klieger, "In Hiring, Algorithms Beat Instinct," *Harvard Business Review*, 2014, *May Print Issue.*

Lakkaraju, Himabindu, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables," in "Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining" KDD '17 Association for Computing Machinery New York, NY, USA 2017, p. 275–284.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, June 2010, *48* (2), 281–355.

Levinson, Harry, "Management by Whose Objectives?," *Harvard Business Review*, 2003, *January.*

Li, Danielle, Lindsey Raymond, and Peter Bergman, "Hiring as Exploration," *Working paper, available online at* https://danielle-li.github.io/assets/docs/HiringAsExploration.pdf, 2021.

Logg, Jennifer M., Julia A. Minson, and Don A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, 2019, *151*, 90–103.

Longoni, Chiara, Andrea Bonezzi, and Carey K Morewedge, "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research*, 05 2019, *46* (4), 629–650.

Mahmud, Hasan, A.K.M. Najmul Islam, Syed Ishtiaque Ahmed, and Kari Smolander, "What influences algorithmic decision-making? A systematic literature review on algorithm aversion," *Technological Forecasting and Social Change*, 2022, *175*, 121390.

Melville, Nigel, Kenneth Kraemer, and Vijay Gurbaxani, "Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value," *MIS Quarterly*, 2004, *28* (2), 283–322.

Oyer, Paul and Scott Schaefer, "Personnel Economics: Hiring and Incentives," Working Paper 15977, National Bureau of Economic Research May 2010.

Pisanelli, Elena, "A new turning point for women: artificial intelligence as a tool for reducing gender discrimination in hiring," *Working paper*, 2022.

_ , "Your resume is your gatekeeper: Automated resume screening as a strategy to reduce gender gaps in hiring," *Economics Letters*, 2022, p. 110892.

Pope, Devin G. and Justin R. Sydnor, "Implementing Anti-discrimination Policies in Statistical Profiling Models," *American Economic Journal: Economic Policy*, August 2011, *3* (3), 206–31.

Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy, "Mitigating bias in algorithmic hiring," in "Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency" ACM jan 2020.

Rajpurkar, Pranav, Jeremy Irvin, Robyn Ball, ..., and Matthew Lungren, " Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Med*, 2018, *15* (11).

Rambachan, Ashesh and Jonathan Roth, "Bias In, Bias Out? Evaluating the Folk Wisdom," in Aaron Roth, ed., *1st Symposium on Foundations of Responsible Computing (FORC 2020)*, Vol. 156 of *Leibniz International Proceedings in Informatics (LIPIcs)* Schloss Dagstuhl–Leibniz-Zentrum für Informatik Dagstuhl, Germany 2020, pp. 6:1–6:15.

_ **, Jon Kleinberg, Jens Ludwig, and Sendhil Mullainathan**, "An Economic Perspective on Algorithmic Fairness," *AEA Papers and Proceedings*, May 2020, *110*, 91–95.

Rock, Daniel, "Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence," *Working paper*, 2021.

Rubin, Donald B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology 66(5): 688-701*, 1974.

Silberzahn, R., E. L. Uhlmann, D. P. Martin, P. Anselmi, F. Aust, E. Awtrey, Š. Bahník, F. Bai, C. Bannard, E. Bonnier, R. Carlsson, F. Cheung, G. Christensen, R. Clay, M. A. Craig, A. Dalla Rosa, L. Dam, M. H. Evans, I. Flores Cervantes, N. Fong, M. Gamez-Djokic, A. Glenz, S. Gordon-McKeon, T. J. Heaton, K. Hederos, M. Heene, A. J. Hofelich Mohr, F. Högden, K. Hui, M. Johannesson, J. Kalodimos, E. Kaszubowski, D. M. Kennedy, R. Lei, T. A. Lindsay, S. Liverani, C. R. Madan, D. Molden, E. Molleman, R. D. Morey, L. B. Mulder, B. R. Nijstad, N. G. Pope, B. Pope, J. M. Prenoveau, F. Rink, E. Robusto, H. Roderique, A. Sandberg, E. Schlüter, F. D. Schönbrodt, M. F. Sherman, S. A. Sommer, K. Sotak, S. Spain, C. Spörlein, T. Stafford, L. Stefanutti, S. Tauber, J. Ullrich, M. Vianello, E.-J. Wagenmakers, M. Witkowiak, S. Yoon, and B. A. Nosek, "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results," *Advances in Methods and Practices in Psychological Science*, 2018, *1* (3), 337–356.

Tambe, Prasanna, "Big Data Investment, Skills, and Firm Value," *Management Science*, 2014, *60* (6), 1452–1469.

_ , Lorin M. Hitt, and Erik Brynjolfsson, "The Extroverted Firm: How External Information Practices Affect Innovation and Productivity," *Management Science*, 2012, *58* (5), 843–859.

_ , Peter Cappelli, and Valery Yakubovich, "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward," *California Management Review*, 2019, *61* (4), 15–42.

Wilson, Christo, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, and Frida Polli, "Building and Auditing Fair Algorithms: A Case Study in Candidate Screening," in "Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency" FAccT '21 Association for Computing Machinery New York, NY, USA 2021, p. 666–677.

Wu, Lynn, Bowen Lou, and Lorin Hitt, "Data Analytics Supports Decentralized Innovation," *Management Science*, 2019, *65* (10), 4863–4877.

Tables

Table 1: **Summary statistics**

	Mean	Std. Dev	Min	Max	N
Female	0.41	0.49	0	1	397
Education					
Undergraduate degree in the US	0.87	0.33	0	1	397
BS degree	0.14	0.34	0	1	397
Top 25 undergraduate institution	0.10	0.30	0	1	397
EMBA student	0.32	0.47	0	1	397
Previous industry					
Finance	0.26	0.44	0	1	397
Technology	0.19	0.40	0	1	397
Business Services	0.17	0.38	0	1	397
Real Estate	0.05	0.22	0	1	397
Services	0.12	0.32	0	1	397
Arts	0.09	0.28	0	1	397
Other	0.12	0.32	0	1	397
Modeling approach used					
Uses cross-tabs	0.13	0.34	0	1	397
Uses regression	0.83	0.38	0	1	397
Uses machine learning	0.04	0.20	0	1	397
Uses sensitive covariates	0.90	0.30	0	1	397
Considers multiple models	0.60	0.49	0	1	397

Notes: This table displays summary statistics at the manager level.

Table 2: **Impact of the hiring algorithm on job performance**

	Job performance		
	(1)	(2)	(3)
Accepted by algorithm	493.5***	493.5***	492.0***
	(25.74)	(25.74)	(26.10)
R^2	0.177	0.177	0.306
Observations	15585	15585	15585
Mean of rejected candidates	2324	2324	2324
Effect size (%)	21.2	21.2	21.2
Participant controls	No	Yes	No
Participant fixed effect	No	No	Yes

Notes: This table examines the impact of adopting the hiring algorithm on job performance. It display the results of a regression of job performance on an indicator for being accepted by the algorithm (relative to rejected candidates). Column 1 includes no controls, column 2 includes manager controls including their gender, education, and prior industry, and programming method used, and column 3 includes manager fixed effects. All regressions include robust standard errors clustered at the manager level.

Table 3: **Impact of the hiring algorithm on gender diversity**

	Female candidate		
	(1)	(2)	(3)
Accepted by algorithm	-0.0284	-0.0284	-0.0275
	(0.0216)	(0.0216)	(0.0219)
R^2	0.001	0.001	0.069
Observations	15585	15585	15585
Mean of rejected candidates	0.470	0.470	0.470
Effect size (%)	-6.1	-6.1	-5.9
Participant controls	No	Yes	No
Participant fixed effect	No	No	Yes

Notes: This table examines the impact of adopting the hiring algorithm on gender diversity. It display the results of a regression of a binary indicator for being a female candidate on an indicator for being accepted by the algorithm (relative to rejected candidates). Column 1 includes no controls, column 2 includes manager controls including their gender, education, and prior industry, and programming method used, and column 3 includes manager fixed effects. All regressions include robust standard errors clustered at the manager level.

Table 4: **Impact of the hiring algorithm on age and region**

Panel A: Age

	Candidate age			
	16-24	25-34	35-49	50-65
	(1)	(2)	(3)	(4)
Accepted by algorithm	0.0763***	0.0732***	-0.0320**	-0.1170***
	(0.010)	(0.010)	(0.013)	(0.010)
R^2	0.104	0.064	0.062	0.071
Observations	15585	15585	15585	15585
Mean of rejected candidates	0.060	0.141	0.504	0.295
Effect size (%)	126.7	52.0	-6.4	-39.8

Panel B: Region

	Candidate region				
	Central & Eastern Europe	Asia	Latin America & the Caribbean	North America & Western Europe	Africa
	(1)	(2)	(3)	(4)	(5)
Accepted by algorithm	0.0033	0.1860***	-0.2000***	0.0143	-0.0042
	(0.011)	(0.016)	(0.014)	(0.017)	(0.003)
R^2	0.062	0.133	0.278	0.132	0.062
Observations	15585	15585	15585	15585	15585
Mean of rejected candidates	0.19	0.12	0.22	0.45	0.01
Effect size (%)	1.8	150.8	-89.1	3.2	-28.6

Notes: This table examines the impact of adopting the hiring algorithm on age and region diversity. It display the results of a regression of a binary indicator for underrepresented groups on an indicator for being accepted by the algorithm (relative to rejected candidates). All regressions include manager fixed effects and robust standard errors clustered at the manager level.

Table 5: **The impact of decreasing female representation on algorithmic adoption recommendations**

	Recommended algorithmic adoption (=1)			
	(1)	(2)	(3)	(4)
Decreased female representation (=1)	-0.287***	-0.293***	-0.290***	-0.290***
	(0.0487)	(0.0485)	(0.0483)	(0.0483)
Impact on job performance, percent		0.0021***	0.0012	0.0012
		(0.001)	(0.001)	(0.001)
Decreased age 50+ representation (=1)			-0.082	-0.082
			(0.056)	(0.056)
Decreased Latin American representation (=1)			0.138**	0.138*
			(0.064)	(0.064)
R^2	0.082	0.096	0.109	0.109
Observations	393	391	391	391
Mean of algorithms with no adverse gender impact	0.670	0.670	0.670	0.670
Effect size (%)	-42.8	-43.6	-43.3	-43.3
Manager controls	No	No	No	Yes

Notes: This table examines the impact of hiring algorithm effects on the decision to adopt. It display the results of a regression of a binary indicator for adopting the algorithm on various diversity and performance measures. All regressions include robust standard errors.

Table 6: **Regression discontinuity design estimates of the impact of decreasing female representation on algorithmic adoption recommendations**

	Recommended algorithmic adoption (=1)			
	(1)	(2)	(3)	(4)
Decreased female representation (=1)	-0.162**	-0.147*	-0.140*	-0.126
	(0.079)	(0.079)	(0.080)	(0.080)
R2	0.025	0.041	0.044	0.089
Observations	167	166	166	166
Mean of rejected candidates	0.670	0.670	0.670	0.670
Effect size (%)	-24.2	-22.0	-20.9	-18.8
Job performance control	No	Yes	Yes	Yes
Other demographic controls	No	No	Yes	Yes
Manager controls	No	No	No	Yes

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines the impact of the gender change of an algorithm on whether it is suggested to be adopted or not. It compares adoption recommendations for hiring algorithms that decrease the number of female hire by one or two, versus algorithms that have no impact or increase the number of female hires by one or two. All regressions include robust standard errors.

Table 7: **The predictors of decreasing female representation**

	Decreased female representation, (=1)		
	(1)	(2)	(3)
Algorithmic approach:			
Uses machine learning	0.0985		0.0770
	(0.173)		(0.176)
Uses regression	0.134		0.104
	(0.0976)		(0.0997)
Excludes sensitive predictors	-0.132**		-0.133**
	(0.0543)		(0.0607)
Tests multiple models	-0.0790		-0.0715
	(0.0502)		(0.0532)
Programmer characteristics:			
Female		-0.0497	-0.0549
		(0.0636)	(0.0668)
EMBA		-0.0385	-0.0419
		(0.0630)	(0.0638)
Has B.S. degree		-0.0513	-0.0445
		(0.0881)	(0.0883)
Went to elite undergrad		-0.00290	-0.00845
		(0.0903)	(0.0914)
Went to undergrad in the US		0.0363	0.0576
		(0.109)	(0.108)
R2	0.021	0.007	0.025
Observations	396	372	371

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines what algorithmic and managerial covariates predict whether a given algorithm would decrease the number of female hires. effects on the decision to adopt. Columns 1 examines what programming techniques influence the likelihood of decreasing the number of female hires, while Column 2 examines how managerial covariates do so. Column 3 includes both. All regressions include robust standard errors.

Table 8: **Complementarities in the impact of HRM practices**

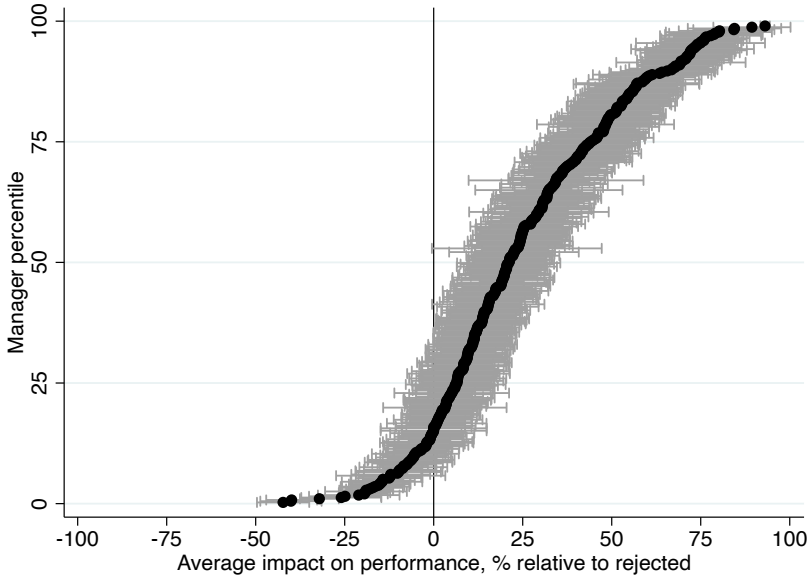
	Performance improvement from algorithm, percent				
	(1)	(2)	(3)	(4)	(5)
Selective hiring	6.105	11.21	4.684		10.57
	(5.787)	(7.582)	(5.695)		(7.901)
Technical hiring	-3.721	0.339		-3.760	-1.973
	(3.952)	(4.200)		(3.748)	(4.281)
Hiring for ML	6.083		4.105	-1.781	-4.610
	(7.232)		(7.552)	(7.796)	(8.022)
Selective X Tech		-16.66			-13.81
		(10.74)			(11.01)
Selective X ML			15.44*		18.08
			(9.350)		(11.16)
Tech X ML				28.62**	30.41**
				(12.56)	(12.90)
Selective X Tech X ML					.
					.
R2	0.006	0.012	0.005	0.008	0.019
Observations	395	395	395	395	395

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines complementarities in the impact of HRM practices on the increase in job performance given by an algorithm using the strategy described in Section 6.2. All regressions include robust standard errors.

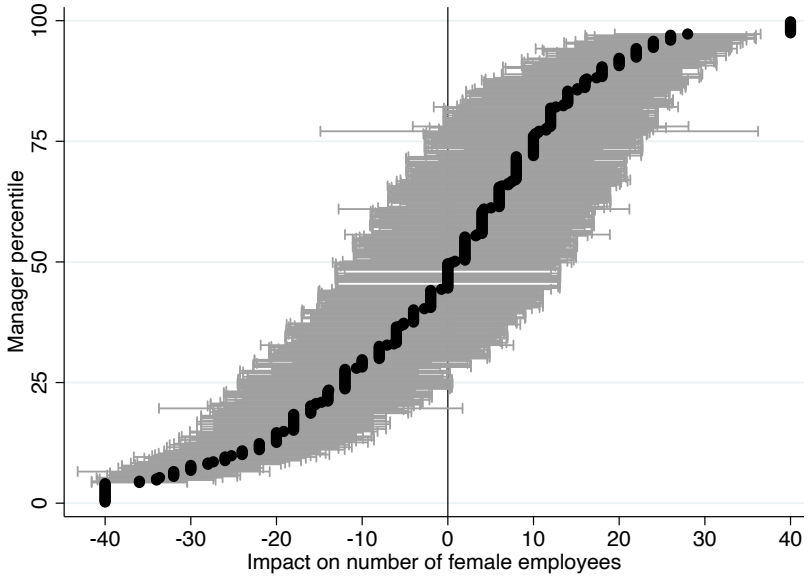
Figures

Figure 1: **Impact of the hiring algorithm on job performance, by manager**



Notes: This figures examines the impact of adopting the hiring algorithm on job performance for each manager. It display the results of a regression of job performance on an indicator for being accepted by the algorithm (relative to rejected candidates), subsetting the sample to each manager. The treatment effects are sorted by largest to smallest.

Figure 2: **Impact of the hiring algorithm on workforce gender, by manager**



Notes: This figures examines the impact of adopting the hiring algorithm on candidate gender for each manager. It display the results of a regression of an indicator of female on an indicator for being accepted by the algorithm (relative to rejected candidates), subsetting the sample to each manager. The treatment effects are sorted by largest to smallest.

Appendix: For Online Publication Only

A Additional empirical results

A.1 Participants by section

Table A.1: Number of participants per section

Section	Date	Type	Number of students
1	Summer 2019	MBA	49
2	Spring 2020	MBA	56
3	Spring 2020	EMBA	18
4	Summer 2020	MBA	50
5	Summer 2020	EMBA	21
6	Summer 2021	MBA	62
7	Summer 2021	EMBA	25
8	Spring 2022	MBA	60
9	Summer 2022	EMBA	56
Total			397

Notes: This table displays the number of students that completed the assignment in each section.

A.1.1 Quantile regression impact on job performance

In Table A.1.1, I examine the impact of the hiring algorithm on candidate job performance using quantile regression.

Table A.2: Quantile regression results for job performance

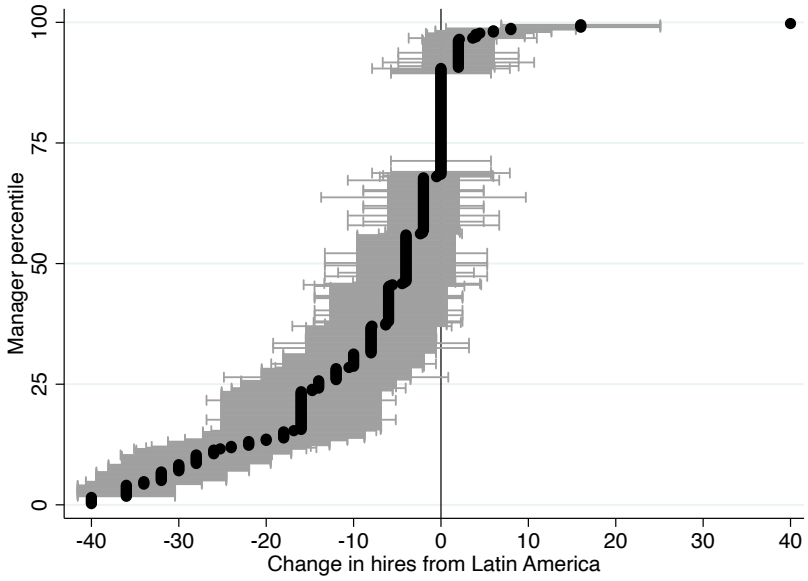
	Job performance				
	Quantile:				
	10	25	50	75	90
	(1)	(2)	(3)	(4)	(5)
Accepted by algorithm	756.1***	650.3***	483.8***	412.2***	320.3***
	(21.39)	(6.58)	(13.30)	(7.45)	(9.91)
Observations	15585	15585	15585	15585	15585
Mean of rejected candidates	1498	1900	2336	2751	3074
Effect size (%)	32.5	28.0	20.8	17.7	13.8

Notes: This table examines the impact of adopting the hiring algorithm on job performance. It display the results of a quantile regression of job performance on an indicator for being accepted by the algorithm (relative to rejected candidates) with manager controls. The quantiles cutoffs are 10, 25, 50, 75, and 90. All regressions include robust standard errors.

A.1.2 Individual-level impacts on demographic characteristics

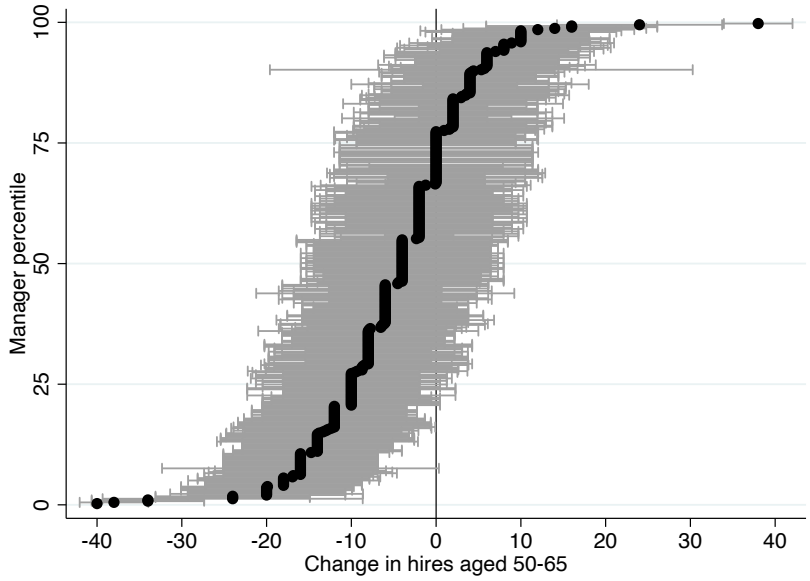
In the main text, I show the individual-level impact of each student's hiring algorithm on firm performance (Figure 1) and the change in female employees (Figure 2). In this section, I reproduce the same analysis except for region (Figure A.1), and age (Figure A.2).

Figure A.1: Impact of the hiring algorithm on number of hires from Latin America , by student



Notes: This figures examines the impact of adopting the hiring algorithm on candidate region for each manager. It display the results of a regression of an indicator for Latin America on an indicator for being accepted by the algorithm (relative to rejected candidates), subsetting the sample to each manager. The treatment effects are sorted by largest to smallest.

Figure A.2: Impact of the hiring algorithm on number of hires aged 50–65 , by student



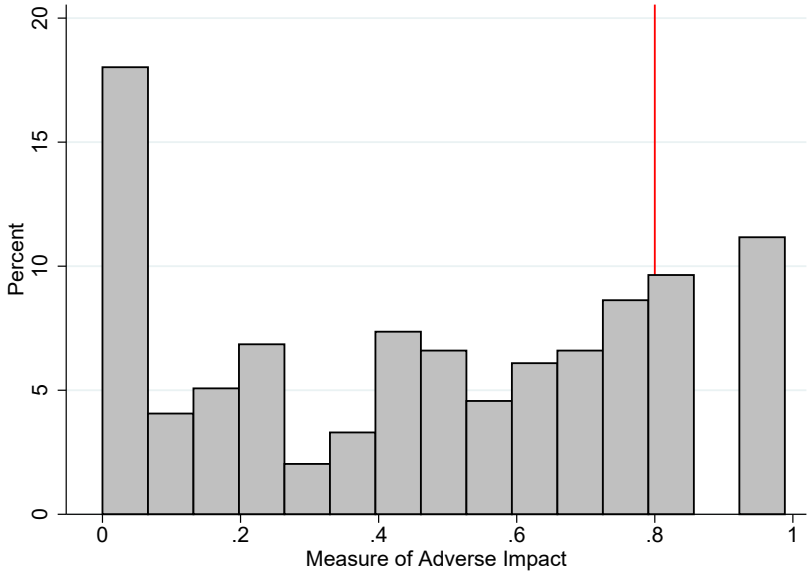
Notes: This figures examines the impact of adopting the hiring algorithm on candidate age for each manager. It display the results of a regression of an indicator for age 50–65 on an indicator for being accepted by the algorithm (relative to rejected candidates), subsetting the sample to each manager. The treatment effects are sorted by largest to smallest.

A.1.3 $\frac{4}{5}^{th}$'s rule analysis

Figure A.3 displays the ratio of selection rates for my adverse impact analysis. I display $\frac{SelectionRate_W}{SelectionRate_M}$ for algorithms where $SelectionRate_W < SelectionRate_M$, and $\frac{SelectionRate_M}{SelectionRate_W}$ for algorithms where $SelectionRate_M \leq SelectionRate_W$. The figure shows that 80 percent of hiring algorithms would fail the $\frac{4}{5}^{th}$'s rule. Out of these, 54 percent had an adverse impact for women, while the remaining 46 percent had an adverse impact for men. Table A.3 displays the number of algorithms in each bucket and their average adoption rate.

Meanwhile, Figure A.4 displays the average recommendation rates for algorithms based on whether they (i) decrease female representation or not, and (ii) would pass or fail the four-fifths test. If fear of illegal behavior were driving my results, I would expect to see that algorithms with an impact ratio of less than 0.80 are less likely to be adapted than those with an impact ratio of 0.80 or more. However,

Figure A.3: Distribution of Ratio of Selection Rates



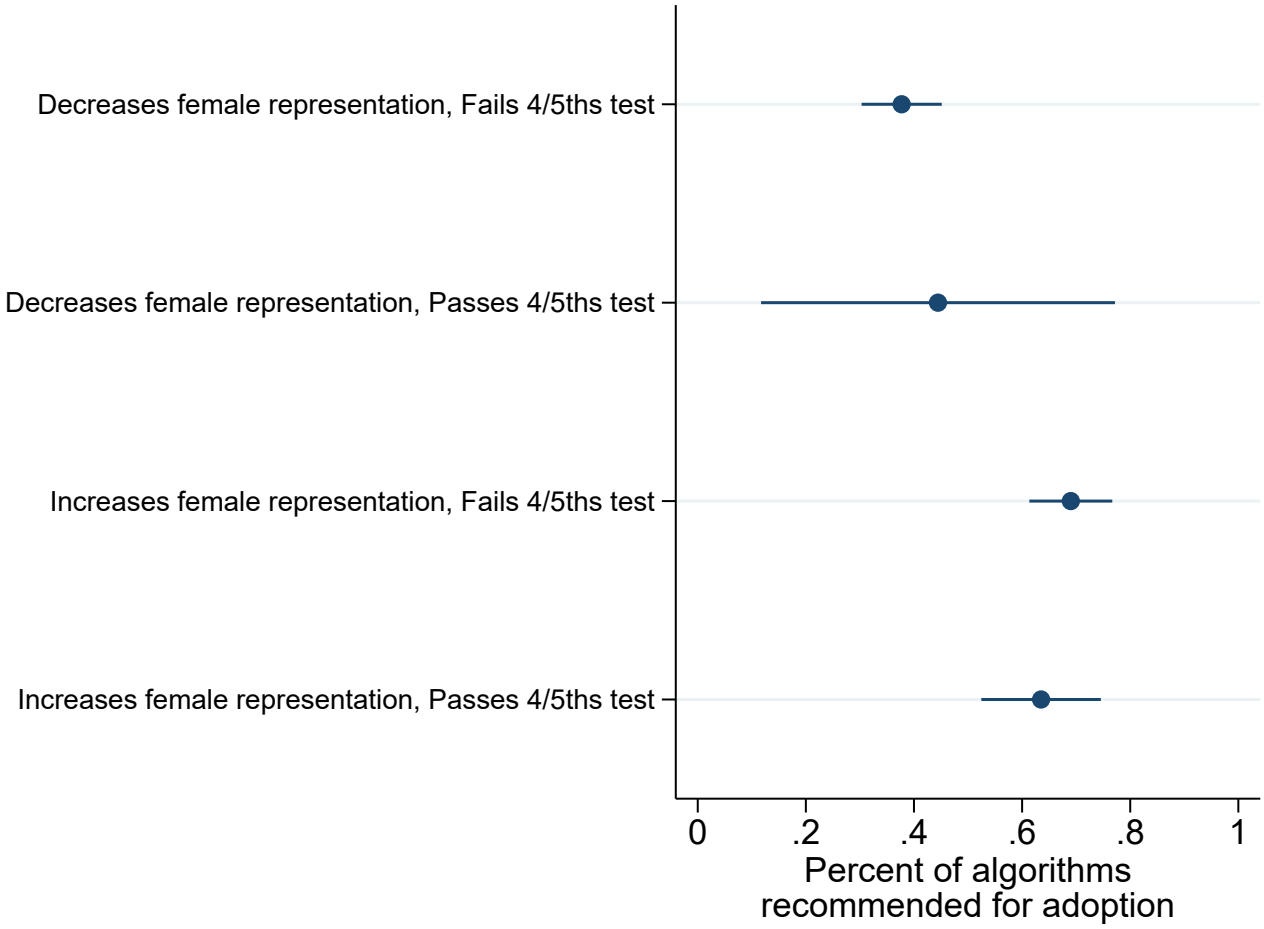
Notes: This figure displays the ratio of selection rates for my adverse impact analysis. I display $\frac{SelectionRate_W}{SelectionRate_M}$ for algorithms where $SelectionRate_W < SelectionRate_M$, and $\frac{SelectionRate_M}{SelectionRate_W}$ for algorithms where $SelectionRate_M \leq SelectionRate_W$.

this is not the case. The results indicate that passing or failing the four-fifths test has no bearing on the likelihood of adoption ($p = 0.69$ and $p = 0.42$ for algorithms that decrease vs increase female representation, respectively). Instead, my results indicate that decision-makers adopt algorithms that increase female representation, regardless of whether they would be legal or not, and avoid algorithms that decrease female representation, regardless of whether they would be legal or not ($p = 0.01$ and $p = 0.28$ for algorithms that fail the four-fifths test, and for algorithms that pass it, respectively). For these reasons, my results are unlikely to be driven by legal fears.

Table A.3: Adoption rates by legality and direction of impact

		Total	Accepted	
		N	N	%
		(1)	(2)	(3)
Failed $\frac{4}{5}^{th}$'s rule	$N_{F,old} > N_{F,new}$	167	63	38
Failed $\frac{4}{5}^{th}$'s rule	$N_{F,old} \leq N_{F,new}$	142	98	69
Passed $\frac{4}{5}^{th}$'s rule	$N_{F,old} > N_{F,new}$	9	4	44
Passed $\frac{4}{5}^{th}$'s rule	$N_{F,old} \leq N_{F,new}$	74	47	64

Figure A.4: Average recommendation rates, by algorithm demographic impact and adverse impact ratio



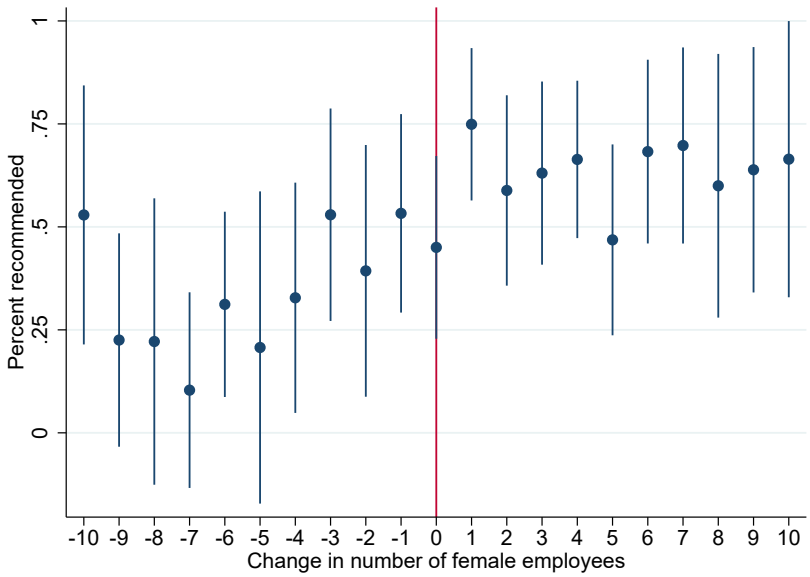
Notes: This figure displays the average adoption recommendation rate for algorithms by whether or not they decrease female representation, and whether or not they would pass the EEOC's four-fifths rule.

A.2 Regression discontinuity design tests

In Section 5.5 of the main text, I show evidence using a regression discontinuity design that marginally decreasing the number of female hires leads to a lower likelihood of algorithmic adoption. The validity of this design depends on algorithms not exhibiting any differences other than their gender impact along the cutoff. In this section, I examine this possibility.

First, I display the average recommendation rate for algorithms based off their impact on gender representation in Figure A.5. This figure provides some visual evidence that algorithms that decrease female representation are less likely to be recommended for adoption. I also present the same figure but using other outcome variables including the job performance impact of the algorithm (Figure A.6), the impact on Latin American hires (Figure A.7), and the impact on the average age of hires (Figure A.8). These figures show no difference in algorithmic impacts to the right versus the left of the gender threshold.

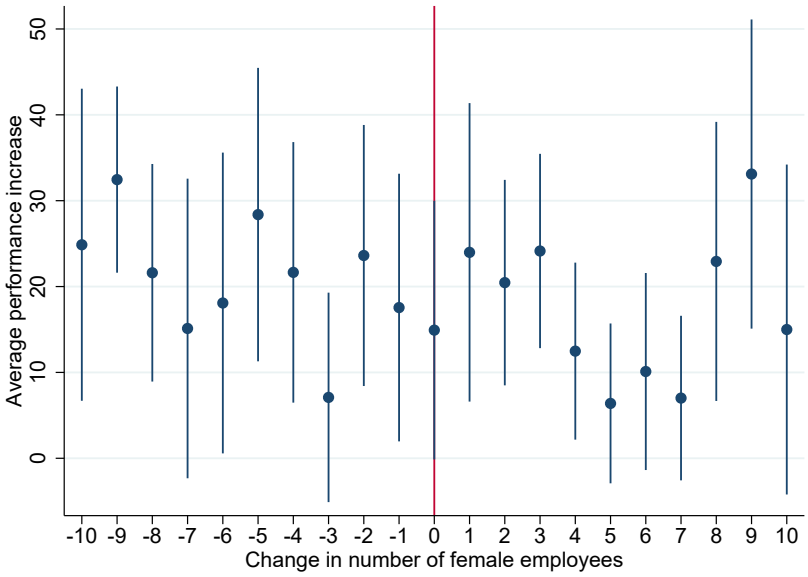
Figure A.5: Average recommendation rates, by impact on female representation ratio



Notes: This figure displays the average adoption recommendation rate for algorithms by their aggregate impact on female representation

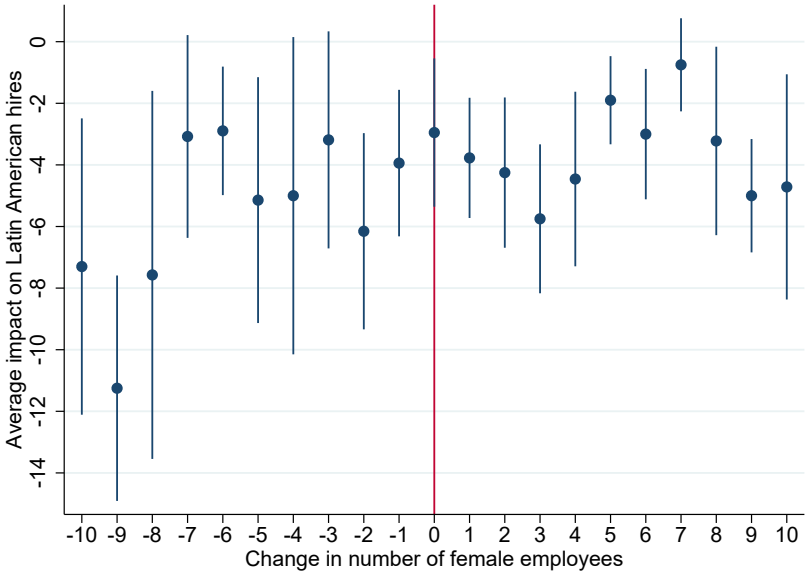
ZThe next test of my regression discontinuity design re-runs the same regression but using placebo outcomes. I run the same regression discontinuity estimate as in Table 6 of the main text, but use various other outcomes as the dependent variable. I display the results in Table A.4. The results show

Figure A.6: Average performance impacts, by impact on female representation ratio



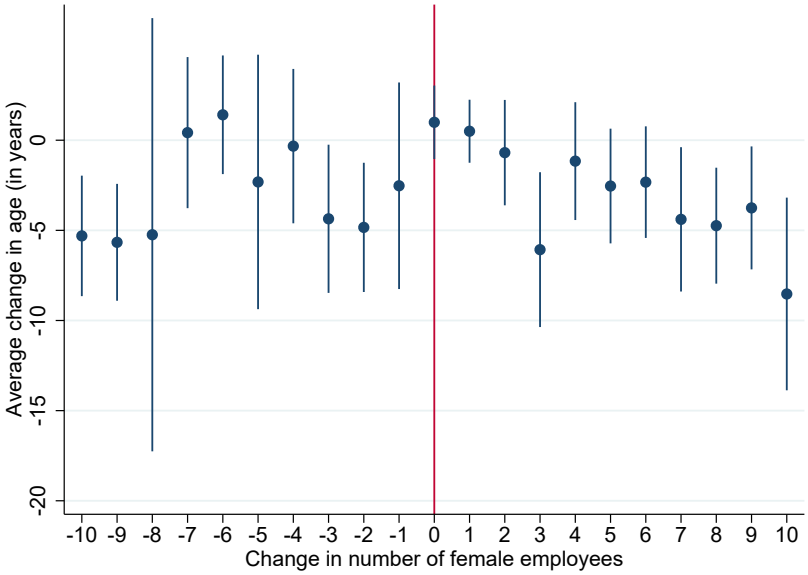
Notes: This figure displays the average performance impact of an algorithm by their aggregate impact on female representation

Figure A.7: Average impacts on the number of Latin American hires, by impact on female representation ratio



Notes: This figure displays the average impact of an algorithm on the number of Latin American hires, by the algorithm's aggregate impact on female representation.

Figure A.8: Average impacts on age, by impact on female representation ratio



Notes: This figure displays the average impact of an algorithm on the average age of hires, by the algorithm's aggregate impact on female representation.

that algorithms to the left versus right of the cutoff do not differ in observable characteristics, further adding credibility to the regression discontinuity design.

I also examine whether the impacts of the regression RDD depend on the comparison group used. More specifically, the regression discontinuity design pools algorithms that had no change in the gender composition of hires with those that marginally increased the number of female hires, as the control group. However, a more specific test of the "gendered loss aversion" hypothesis would compare adoption recommendations for algorithms that weakly increase versus weakly decrease female hires relative to the control group of no change. I display these results in Table A.5. The results indicate that algorithms that weakly increase the number of female hires were just as likely to be adopted as those that had no impact ($p = 0.33$) but much less likely to be adapted than those that increased female representation ($p = 0.01$).

Table A.4: **Regression discontinuity design estimates of other (placebo) outcomes**

	Impact on:		
	Decreased hires aged 50-65, (=1)	Decreased Latin American hires, (=1)	Change in job performance, percent
	(1)	(2)	(3)
Decreased female count (=1)	0.121	-0.0656	-6.059
	(0.0765)	(0.0777)	(4.652)
R2	0.014	0.004	0.010
Observations	167	167	166

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines the impact of the gender change of an algorithm on various placebo outcomes. It compares placebo outcomes for hiring algorithms that decrease the number of female hires by one or two, versus algorithms that have no impact or increase the number of female hires by one or two. All regressions include robust standard errors.

Table A.5: **Regression discontinuity design estimates, using no change as the control group**

	Recommended algorithmic adoption (=1)			
	(1)	(2)	(3)	(4)
Decreased female count (=1)	-0.105	-0.0773	-0.0747	-0.0386
	(0.137)	(0.134)	(0.140)	(0.141)
Increased female count (=1)	0.112	0.125	0.125	0.137
	(0.130)	(0.126)	(0.133)	(0.134)
R2	0.043	0.058	0.058	0.100
Observations	167	166	166	166
P-values:				
Decrease vs no change	0.445	0.564	0.593	0.784
Increase vs no change	0.391	0.322	0.348	0.306
Decrease vs increase	0.008	0.014	0.017	0.039

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Notes: This table examines the impact of the gender change of an algorithm on whether it is suggested to be adopted or not. It compares adoption recommendations for hiring algorithms that decrease or decrease the number of female hires by one or two, versus algorithms that have no impact on the number of female hires. All regressions include robust standard errors.

B Task Instructions

B.1 Background

On Canvas, you will find Excel data containing the personal characteristics of about 2,000 workers who are eligible to be hired by a company. Note: This is a real dataset. However, we have anonymized the company and refer to it as "ACME Corp." ACME is a large diversified multinational who extensively utilizes strategy and management consultants and often hires their consultants into headquarters staff and executive roles. ACME employs mostly mid-skill workers in developed countries. Their jobs require intermediate math literacy. Suppose that you're a recent MBA graduate asked to help optimize ACME's employment practices. You can think of yourself either as a junior executive at ACME, or as one of the consultants hired by ACME. For your homework assignment, analyze the data and answer the questions below. This is a very small scale-exercise. In the real world, your data will probably include more than 2,000 candidates and more than 14 variables per candidate. The goal of this exercise is to give you experience working on the underlying concepts in larger-scale analysis of strategy (and discussed in class).

B.2 Part One

Answer Part 1 questions 1-3 using up to 300 words for each question. Defend your answers with analysis of this data. You may utilize tables and/or charts as part of your explanation.

- What variables are the best predictors of being hired by ACME?
- What variables are the best predictors of job performance for workers who were hired?
- Formulate a hypothesis about how ACME could improve its hiring. Justify your reasoning.[1]
- Using your hypothesis above, list the CandidateIds of:
  - The top 20 rejected workers ACME should have hired under your hypothesis.
  - The 20 hired workers ACME should have rejected.

[1] Your dataset does not include estimates of how much each worker would cost to hire (i.e., their salary requirements) or their probability of accepting an offer. Obviously this would be an important component to a hiring strategy. For the purposes of this assignment, assume that workers demand the same salary and be equally likely to accept the offer. That is: Focus on the worker-quality issues, and assume that cost considerations will be someone else's job later.

Upload your answers to Part 1 on Canvas. After you do, we'll share more data with you about the job performance of the 20 CandidateIds you listed in 4a.

B.3 Part Two

Using the new data emailed to you, evaluate your hypothesis in Part 1, Questions 3 and 4. Answer all Part 2 questions using up to 300 words for each question. Defend your answers with analysis of this data. You may utilize tables and/or charts as part of your explanation.

- Are your 20 proposed hires better or worse than the 20 you rejected?
- Did your make ACME's workforce more diverse? Or better in other ways that aren't captured by job performance?
- Combining your answers to Part 2, Q1 and Q2. Assume there are no alternatives besides your proposal and continuing with the status quo. Do you recommend that ACME adopt your proposal?