July 2024

Working Paper

Perceptions of algorithmic criteria: The role of procedural fairness

Lily Morse, Mike H. M. Teodorescu, and Yakov Bart

This working paper is available online at: https://www.brookings.edu/center/center-on-regulation-and-markets/

Center on Regulation and Markets at BROOKINGS

The Center on Regulation and Markets at Brookings creates and promotes rigorous economic scholarship to inform regulatory policymaking, the regulatory process, and the efficient and equitable functioning of economic markets. The Center provides independent, non-partisan research on regulatory policy, applied broadly across microeconomic fields.

Perceptions of Algorithmic Criteria: The Role of Procedural Fairness[1]

Working Paper for Brookings Working Paper Series

Lily Morse*, Mike H. M. Teodorescu**, and Yakov Bart***

*College of Business and Economics, West Virginia University, Morgantown WV.

** Information School, University of Washington, Seattle WA,

Responsibility in AI Systems & Experiences (RAISE) Center, University of Washington.

*** D'Amore-McKim School of Business, Northeastern University, Boston MA.

Abstract

The rise of artificial intelligence (AI) has enabled modern society to automate aspects of the organizational hiring process. Yet, prospective job candidates are hesitant to engage with such technologies in their everyday lives unless they perceive algorithms as behaving fairly. Procedural fairness is considered critical in shaping individual attitudes toward algorithms. However, empirical studies examining the role of procedural fairness in AI-enabled hiring are lacking. The present research seeks to bridge this gap by investigating how perceptions of procedural fairness and related fairness dimensions influence job applicants' perceptions of different hiring algorithms designed to incorporate fairness ideals and their attitudes toward companies using these algorithms. Our findings indicate that people perceive hiring algorithms as procedurally fairest when they adopt a "Fairness through unawareness" approach to mitigating bias. They are also likely to view companies who use this approach more positively and are more motivated to apply for open positions.

[1] This research has been supported by the Northeastern University TIER 1 Seed Grant program.

Introduction

The growing use of algorithms to predict organizational performance and other metrics provides opportunities to streamline workplace inefficiencies while increasing reliance on data-driven decisions that minimize human errors and biases. However, the literature on algorithms indicates that public knowledge of them is limited and that people may not consider them particularly important to major life decisions, such as whether they are hired or not. In fact, many believe that algorithms could do more harm than good and are inherently biased.

Individuals may distrust algorithms because they are impersonal, lack nuance, invade privacy, and/or are unfair. The field of fairness in machine learning is relatively recent and strives to help bridge that adoptability gap by providing statistical criteria on whether an algorithm is fair in its predictions (e.g., whether a prospective job candidate would be a good fit with a particular company) with regards to certain individual attributes widely considered by law to be protected. Protected attributes include demographics such as race, age, religion, ethnicity, socioeconomic status, and gender. Fairness cannot be easily defined, as it is a concept that has been at the intersection of philosophy, law, and computer science. For example, Mehrabi et al. (2021, p. 1) consider *fairness* as preventing "prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics". Our focus in this paper is on a more nuanced approach to evaluating fairness based on the literature on organizational justice, with an emphasis on procedural fairness, as detailed in Morse and colleagues (2022).

Procedural fairness requires "fairness of the decision-making procedures" (Colquitt and Rodell, 2011), originally defined outside of the algorithmic context, which involves evaluating decision processes on six components, specifically consistency, accuracy, ethicality, representativeness, bias suppression, and correctability (Leventhal, 1980). Some of these

components are defined in laws such as the new EU AI Act[2] or the New York City Local Law 144, which deals specifically with fairness under algorithmic hiring.[3] Correctability, for example, is related to the ability of users to request the replacement of incorrect data that was used in the algorithm decisions (Teodorescu & Makridis, 2024) and requires the ability of audit committees to demand changes to algorithms if errors are found after the algorithm-based system is launched to the public (Tarafdar et al., 2020).

Although fairness cannot easily be defined, researchers have attempted to capture its meaning by breaking it into several aspects or by proposing distinct models of fairness. Public ideas about the causes of algorithmic unfairness may not align with the causes proposed by computer scientists, though more research is needed. A person's views on an algorithm appear to depend on the decision-making scenario, but it is unclear whether perception differs among demographics or is affected by the explanation of the decision. An experiment with users of diverse backgrounds exposed to popular fairness criteria may help promote our understanding of how to design fairness criteria that are explainable (within the broader context of explainability of AI, see Confalonieri et al., 2021) and more usable by users without a background in AI.

In the next sections, we provide a brief overview of the results from the empirical study we conducted and review the literature on perceptions toward algorithms.

Summary of Results

We conducted a between-subjects online behavioral experiment that sought to understand how prospective job applicants perceive the use of algorithmic fairness criteria in the hiring context. Specifically, we assessed whether three popular algorithmic fairness criteria could help

[2] https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
[3] https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

to improve individuals' fairness judgments, especially procedural fairness, compared to a Control condition (in which algorithms contained no specified fairness approach). We find that people viewed algorithmic criteria that were explicitly blind to job applicants' demographic characteristics (e.g., age, gender, race) more positively than the Control: perceptions of procedural fairness along with other perceived fairness measures were all higher for this fairness criterion. These fairness perceptions played a positive role in attracting people to companies that used this approach. The same cannot be said for the other algorithmic fairness criteria we tested.

Literature Review: Perceptions of Algorithms

Perception of algorithms by the public

Previous research suggests that much of the public has a limited understanding of algorithms and may not consider them very impactful or important. Indeed, an entire substream of literature in this space, dubbed "algorithmic aversion", focuses on understanding why people have a generalized dislike toward algorithms (Jussupow et al., 2020; Logg et al., 2019). In a survey by Araujo et al. (2018), most participants said they knew little about algorithms or artificial intelligence, with 45% reporting that they knew nothing at all about algorithms. When asked participants to rate the importance of automated decision-making (ADM) and other societal concerns such as the environment and safety/crime, ADM was deemed important by about 35% of subjects, whereas the other concerns were all deemed important by over 70% of subjects.

Even if the public does not regard algorithms as highly important, there appear to be significant concerns about their usage. Less than a quarter of participants in Araujo et al. (2018) thought ADM would make life better, while over half thought it could be detrimental. Across seven studies, Reich and colleagues (2022) similarly found that people are hesitant to use

algorithms because they believe—often incorrectly—that algorithms cannot learn from their mistakes. Perhaps the broadest reason for apprehension is perceived unfairness, which can arguably include concerns such as inaccuracy and bias. Algorithmic unfairness has thus been the focus of a substantial amount of research.

Perceptions of algorithmic fairness

The literature suggests that fairness, as humans perceive it, cannot be simplified into a single definition. Though algorithms can be mathematically "proven" as fair, subjects may still view an algorithm's decisions as less than fair (Koenecke et al., 2020). For instance, Newman et al. (2020) showed that people considered algorithms designed to make personnel decisions to be procedurally unfair, which was driven by perceptions that algorithms lacked the ability to make holistic judgments about the people being evaluated. This finding corroborates prior work by Smith (2018, p. 1) which showed that people believe algorithms "remove the human element from important decisions". More specifically, survey results revealed that 54% of respondents believed a personal finance score algorithm would effectively find good customers, but only 32% thought it would be fair to the persons being analyzed. These disparities are the crux of this paper: prior studies provide support that theoretically fair, non-discriminatory, and effective algorithms are not necessarily perceived as fair.

Attempts to define algorithmic fairness in the computer science literature have generally taken one of two approaches: proposing aspects to measure fairness or comparing different models of fairness. In the first category, for example, Grgić-Hlača et al. (2018) proposed eight latent properties of features. The authors hypothesized that a person's judgment of a feature's latent properties could explain their judgment of the feature's fairness. For example, two of the properties were *relevance* and *causes outcome*. Someone who thought a feature was irrelevant to

a decision might judge it as unfair to use, and someone who thought a feature could cause a certain outcome might judge it as fair to use. Similarly, Binns and colleagues (2018) developed five justice constructs: *agreement with the decision*, *understanding of the decision-making process*, *appropriateness of factors used*, *fairness of the process*, and *whether the individual deserved the outcome*.

While these studies treated fairness as a combination of factors, other work has compared decision-making models that could each be considered fair. To fairly divide something desirable among individuals, Saxena et al. (2019) considered three approaches: treating similar individuals similarly, never favoring a worse individual over a better one, and calibrated fairness (i.e., choosing individuals in proportion to their merit). Lee and colleagues (2017) also considered the scenario of distributing a desirable resource, specifically food donations among non-profits. However, their models were based on different definitions of fairness—*equality* and *equity*—and included a third model based on *efficiency* that simply selected the non-profit closest to the donation pick-up site.

A separate yet related stream of research in the organizational and psychological literatures has focused on people's attitudes and views toward fairness. As mentioned before, there are many definitions of fairness which together broadly focus on perceptions of fairness with respect to decision outcomes, decision-making procedures, and the quality of their interactions with decision-makers in organizations (Colquitt, 2012). These definitions are better known as distributive, procedural, informational, and interactional fairness. With respect to algorithms, computer science perspectives toward fairness tend to prioritize distributive fairness in that they seek to ensure that algorithms provide a clear fair/unfair outcome (Robert et al., 2020). Notably, this work tends to overlook procedural fairness elements (i.e., decision attributes

that inform how decisions are made) despite findings indicating that procedural fairness is more influential than distributive fairness in shaping individuals' fairness judgments (Morse et al., 2022; van den Bos et al., 2001).

Effects of different individual and situational factors

Various studies have investigated how characteristics of the decision-making scenario, of the participants in the research, and the explanation of decisions affect user perceptions of algorithms.

One finding reflected in multiple experiments is that opinions depend on the type of decision being made and the people it involves (Dastile et al., 2020; Raghavan et al., 2020). For instance, when Smith (2018) proposed four different scenarios, the proportion of respondents that found each acceptable differed by up to 11%. Lee (2018) found that when evaluating employee performance and hiring, people trusted a human's decision more than an algorithm's for tasks involving human skills. For tasks involving mechanical skills, they trusted the decisions equally.

Paul and Ahmed (2023) demonstrated that gender and age influenced people's perceptions of the fairness and efficacy of algorithms. When the decision impacted the individuals being evaluated, the outcome of the decision influenced their opinion. Wang et al. (2020) similarly found that individuals whom an algorithm judged favorably viewed it as fairer. Results also revealed that computer literacy — but not age, gender, level of education, or race — significantly affected people's judgments of an algorithm's fairness, suggesting that a person's expertise with machines is an important predictor.

When van Berkel and colleagues (2019) asked whether a recidivism-predicting algorithm should use certain demographics, respondents' demographic characteristics did not play a role:

e.g., a person's gender did not affect whether they thought gender should be considered. When Scurich and Krauss (2020) asked Californians about a bill that would replace cash bail with an RAI (Risk Assessment Instruments, commonly used in criminal justice algorithmic approaches), they did not find a significant association between support for the legislation and demographic factors. However, the same study found that 76% of participants of color, compared to 63% of White participants, thought that RAIs would exacerbate the criminal justice system's racial inequalities. Smith (2018) also found differences across demographic groups. About one-third of adults 50 and older, compared to half of 18- to 29-year-olds, believed that computer programs can make unbiased decisions. Twenty-five percent of White respondents, compared to 45% of Black respondents, thought it was fair to judge a person's financial responsibility with an algorithm that used data about personal characteristics and behaviors. On the other hand, 49% of White respondents, compared to 61% of Black respondents, thought it was unfair to judge a person's eligibility for parole in a similar way.

Finally, Binns et al. (2018) investigated whether the way a decision was explained affected individual attitudes toward algorithms. When participants were exposed to multiple explanation styles, case-based explanations (i.e., explanations that used a similar case from the model's data to justify the decision) appeared to have the most negative impact on fairness perceptions. When participants were exposed to only one style, explanation styles did not significantly impact fairness perceptions.

In the following section, we review scholarship on algorithms used in the hiring and performance evaluation context.

Literature Review: Algorithms in Hiring and Performance Evaluation Contexts

Hiring and Performance Evaluation

Hiring decisions affect personal well-being and identity, as many individuals spend more than half of their adult waking hours performing work and developing ties to their careers that become part of their identity (Wrzesniewski et al., 1997). Given that so much of people's lives are spent at work, fairness in hiring algorithms is highly significant to an individual's well-being, as it can affect their socioeconomic status and long-term career prospects. Thus, we chose hiring as the context for the current study.

Productivity improvement is the main reason employers are interested in adopting AI tools, as for other industries. According to the McKinsey Global Institute (2023), AI could add between $60 billion and $90 billion in value in the talent management/HR sector. Hiring, especially in technical positions, is a difficult and resource-consuming process for an organization (Rudman et al., 2016). Manual resume screening by HR employees is insufficient as it does not necessarily translate well into how skills displayed on paper will yield performance in the workplace (Kokkodis et al., 2015). Firms that have R&D jobs or manufacturing jobs often need to test the skills of prospective hires by using their technically trained staff and taking them away from production and other revenue-generating tasks. In other words, hiring is a costly organizational process.

The advent of machine learning tools that assess the desirability of job candidates through resume analysis (Cowgill, 2020) or automated interviewing tools, such as HireVue or AMCAT, can significantly reduce organizational costs.[4] However, it can also cause unfavorable reactions in those who are being evaluated (Langer et al., 2020; Teodorescu et al., 2021). For

[4] https://www.hirevue.com/ and https://www.myamcat.com/.

instance, a recent study found that people's attitudes toward training sets for automated video interviewing tools differed depending on their country of residence (Teodorescu et al., 2022). Since hiring is a particularly challenging environment for algorithm adoption (Lavanchy et al., 2023) and is a recent area of interest from policymakers, being subject to regulation such as the new EU AI Act or the New York City Law 144, we chose it for our experimental context, detailed in the next section.

Methods and Experimental Context

Fairness Criteria Being Tested and Hypotheses

The literature on machine learning fairness has identified over 20 different fairness criteria, mostly based on statistical properties based on protected attributes (Mehrabi et al., 2021). Most of these criteria fall under "group fairness", specifically that the predictive outcomes or some characteristics of the prediction have to be close or identical by groups of the protected attributes. "Demographic parity", by definition, implies that the predicted decision is independent of the protected attribute (Veale & Binns, 2017). For example, the prediction of whether someone would perform well on a job task should not depend on race or gender. The "Equalized odds" fairness criterion implies the equality of the algorithm's True Positive Rates (TPR) and False Positive Rates (FPR) over the protected attributes in the data (Hardt et al., 2016). However, since this condition is rarely satisfied in practice (see Teodorescu & Yao, 2021), there is a restricted version of "Equalized odds" called "Equalized opportunity", which matches only the True Positive Rates and is commonly considered as a better criterion than "Demographic parity" (Hardt et al., 2016). "Fairness through unawareness" (Kusner et al., 2017) refers to leaving out the protected attributes from the data used in the prediction algorithm entirely, making it impossible in theory for the algorithm to discriminate based on these

attributes. In practice, this is often false (Awwad et al., 2020), as there can be redundant encodings, i.e., the same information as protected attributes can appear across other variables that are not explicitly considered to be protected attributes.

In the related theoretical work of some of the authors (Morse et al., 2022), we conceptually argue that the hiring context represents a setting in which concerns about bias and unfair treatment are expected to be highly salient to job applicants and related stakeholders. Therefore, technical fairness criteria that remove barriers to entry for members of historically disadvantaged groups are more likely to be viewed as procedurally fair. Of the three criteria we focus on in our experiment, Morse and colleagues (2022) theorize that "Fairness through unawareness" has the lowest capacity to signal procedural fairness, followed by "Demographic parity", and, finally, "Equalized opportunity". However, the question of how individuals interpret these criteria in practice remains open. Ultimately, if the procedural aspects of a particular criterion are not discernable to the general population, while theoretically fairer, such criteria may not be necessarily better.

Our hypotheses are:

H1: *Algorithms with fairness constraints will be perceived as procedurally fairer than algorithms that are not explicitly designed for fairness.*

H2: Greater salience of an algorithm's procedural fairness will lead to more positive attitudes toward companies using such algorithms.

H3: Procedural fairness perceptions will mediate the relationship between algorithmic fairness criteria and company attractiveness and trustworthiness judgments.

These hypotheses are directly related to the theory section of prior theoretical work (Morse et al., 2022), which has yet to be tested empirically. To provide a comprehensive

investigation of perceived fairness, we also examine distributive, informational, and overall fairness for exploratory purposes.

Participants

A total of 575 individuals participated in the study on Amazon Mechanical Turk. To be eligible, participants were required to be currently employed full-time, 18 years or older, and based in the United States. They were paid $1.25 for their participation.

We pre-registered the study (as well as prior pilots) on Wharton's Credibility Lab Online Registration Platform "As Predicted".[5] We originally aimed to collect data from 600 participants; we ended up with 575 after excluding 25 people for missing a pre-registered attention check.[6]

Design

Participants were randomly assigned to one of four algorithmic fairness conditions: "Fairness through unawareness", "Demographic parity", "Equal opportunity", or Control. The Control condition had no algorithmic fairness constraints in its description.

Procedure

The online survey began with a hypothetical job search scenario. Participants imagined they were looking for a new job at a company in their field and read a job posting they thought could be a good fit. However, they noticed in the posting that the company uses artificial intelligence algorithms to evaluate applicants. All conditions contained the following information about the algorithms:

---

[5] Registration for the study under AsPredicted ID # 156825; pilot registration under ID # 141203.

[6] The attention check was designed to prevent bots and inattentive respondents from completing the study ("Combing your hair is a very important behavior. When answering the question below, please select "never". Combing your hair every day helps discourage knotting and promote follicle growth").

NOTICE OF USE OF ARTIFICIAL INTELLIGENCE ALGORITHMS

> *"Our applicant screening process is partly automated by artificial intelligence algorithms. Algorithmic screening allows us to provide a more accurate and efficient hiring experience by processing job applications quickly and narrowing down to a pool of qualified candidates who best match the requirements of open positions. The algorithms are designed to uphold inclusive and fair recruiting practices with respect to race, color, gender, age, religion, disability, nationality, sexual orientation, or other legally protected characteristics."*

In the "Fairness through unawareness" condition, the notice continued:

> *"The algorithms do not consider applicants for employment based upon these characteristics. By remaining blind, the algorithms ensure that applicants are treated equally, which we believe is central to the success of our organization."*

In the "Demographic parity" condition:

> *"The algorithms continually screen for disparities based on these characteristics to ensure equal outcomes, which we believe is central to the success of our organization. As an example, the algorithms ensure that all black and white applicants have similar rates of being selected for an interview."*

In the "Equal opportunity" condition:

> *"The algorithms continually screen for disparities based on these characteristics in order to ensure equal opportunities, which we believe is central to the success of our organization. As an example, the algorithms ensure that black and white applicants who are qualified for the position have similar rates of being selected for an interview."*

No additional information was provided in the Control condition.

All participants answered two comprehension check questions that they were required to answer correctly to proceed in the study. They also answered manipulation check questions after reading the scenario. Next, participants were asked a randomized series of questions assessing the perceived procedural, distributive, and informational fairness of the algorithms described in

Figure 1. Items underlying the fairness measures (procedural, distributive, informational, and overall fairness).

	Strongly Disagree (1)	Disagree (2)	Neither Agree Nor Disagree (3)	Agree (4)	Strongly Agree (5)
The procedures are applied consistently. (1)	○	○	○	○	○
The procedures are free of bias. (2)	○	○	○	○	○
The procedures are based upon accurate information. (3)	○	○	○	○	○
I am able to appeal the potential outcomes arrived at by the procedures. (4)	○	○	○	○	○
The procedures uphold ethical and moral standards. (5)	○	○	○	○	○
The procedures are explained thoroughly. (6)	○	○	○	○	○
The explanations regarding the procedures are reasonable. (7)	○	○	○	○	○
The notice about algorithms was candid in communications with me. (8)	○	○	○	○	○
The outcomes of this process are fair to job seekers. (9)	○	○	○	○	○
The decisions that the organization makes as a result of this process will be fair. (10)	○	○	○	○	○
The algorithms will lead the organization to make great hiring decisions. (11)	○	○	○	○	○
To show that you are paying attention, please select 'Strongly Disagree'. (16)	○	○	○	○	○

the scenario (see Figure 1). Then, participants reported their perceived overall fairness of the algorithms.

The next portion of the survey assessed perceptions of the company that made the job posting. Participants were presented with a randomized series of questions assessing the company's perceived attractiveness and trustworthiness. Next, participants indicated the likelihood they would apply for the open job position. The final portion of the study asked demographic and occupational questions as well as questions about participants' views toward AI and whether recent legislative or judicial rulings (i.e., regulation) influenced their responses to questions presented earlier in the survey.

Measures

Unless otherwise noted, all questions used 5-point scales: 1 = *Strongly disagree*, 5 = *Strongly agree*.

Procedural Fairness. Participants completed a five-item procedural fairness scale adapted from Colquitt (2001).

Distributive Fairness. Participants completed a three-item distributive fairness scale adapted from Colquitt (2001).

Informational Fairness. Participants completed a three-item informational fairness scale adapted from Colquitt (2001).

Overall Fairness. Participants completed a four-item fairness scale developed by Conlon et al. (2004).

Company Attractiveness. Participants completed a five-item scale used to assess perceptions of employer attractiveness (Highhouse et al., 2003). An example item is "For me, this company would be a good place to work."

Company Trustworthiness. Participants completed an eight-item trust scale often used to assess trustworthiness (McKnight et al., 2002). An example item is "The company has my interests in mind."

Apply Decision. Participants indicated how interested they were in applying for the open job position at the company (1 = *Not at all*, 5 = *Very Much*).

Control variables. We controlled for participant gender, race, age, career tenure (# of years), political beliefs (1 = *Very liberal*, 7 = *Very conservative*), and views toward affirmative action (1 = *Completely illegal*, 5 = *Completely legal*). We also included five items assessing participants' general views toward AI (e.g., "I think AI can be used to benefit people in the workplace"; 1 = *Strongly Disagree*, 5 = *Strongly Agree*).

Data Coding and Descriptives

To test our hypotheses, we created three dummy-coded variables to examine the effects of the experimental conditions against the control condition. This coding approach allowed us to understand the influence of an algorithmic fairness criterion compared to an algorithm without an explicit fairness design. Specifically, the *"*Fairness through unawareness*"* condition was coded as 1 = Fairness through unawareness condition, 0 = Demographic parity condition, 0 = Equal opportunity condition, -1 = Control condition; the *"*Demographic parity*"* condition was coded as 1 = Demographic parity condition, 0 = Fairness through unawareness condition, 0 = Equal opportunity condition, -1 = Control condition; the *"*Equal opportunity*"* condition was coded as 1 = Equal opportunity condition, 0 = Fairness through unawareness condition, 0 = Demographic parity condition, -1 = Control condition.

Descriptive statistics, internal consistency reliabilities, and bivariate correlations are provided in Tables 1 and 2.

Table 1. ***Descriptive Statistics for Criterion Variables by Condition***

	Condition			
Criterion Variables	**Control**[a] M (*SD*)	**Fairness through unawareness**[b] M (*SD*)	**Demographic parity**[c] M (*SD*)	**Equal opportunity**[d] M (*SD*)
Procedural Fairness	3.35 (*0.72*)	3.57 (*0.70*)	3.31 (*0.77*)	3.39 (*0.78*)
Informational Fairness	3.79 (*0.65*)	3.84 (*0.78*)	3.76 (*0.74*)	3.87 (*0.67*)
Distributive Fairness	3.37 (*0.92*)	3.57 (*0.94*)	3.31 (*0.98*)	3.39 (*0.99*)
Overall Fairness	3.43 (*0.96*)	3.66 (*1.04*)	3.66 (*1.06*)	3.45 (*1.07*)
Employer Attractiveness	3.34 (*0.97*)	3.40 (*1.04*)	3.19 (*1.06*)	3.32 (*1.10*)
Employer Trust	3.25 (*0.76*)	3.30 (*0.88*)	3.14 (*0.81*)	3.25 (*0.84*)
Apply	2.86 (*1.24*)	2.97 (*1.28*)	2.76 (*1.26*)	2.81 (*1.28*)
Decision				

[a]$n = 138$, [b]$n = 140$, [c]$n = 139$, [d]$n = 140$

Table 2. ***Bivariate Correlations and Descriptive Statistics***

Variable	M (*SD*)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Fairness through unawareness	0.00 (0.71)	--															
2. Demographic parity	0.04 (0.71)	.50***	--														
3. Equal opportunity	0.00 (0.71)	.50***	.50***	--													
4. Procedural Fairness	3.40 (*0.75*)	.10*	-.03	.02	(.72)												
5. Informational Fairness	3.81 (*0.71*)	.03	-.02	.04	.66***	(.75)											
6. Distributive Fairness	3.41 (*0.96*)	.07†	-.02	.01	.84***	.65***	(.77)										
7. Overall Fairness	3.48 (*1.04*)	.08†	-.02	00	.81***	.65***	.89***	(.97)									
8. Employer Attractiveness	3.31 (*1.04*)	.02	-.05	-.01	.70***	.57***	.75***	.76***	(.94)								
9. Employer Trustworthiness	3.24 (*0.83*)	.02	-.05	.00	.79***	.62***	.79***	.79***	.77***	(.93)							
10. Apply Decision	2.85 (*1.26*)	.03	-.03	-.01	.61***	.50***	.67***	.68***	.83***	.69***	--						
11. AI Views	3.13 (*0.85*)	-.02	-.06	.04	.58***	.46***	.61***	.62***	.63***	.66***	.60***	(.88)					
12. Regulation	3.33 (*1.23*)	.00	-.07	.02	.04	.03	.01	.03	-.01	.03	.03	.09*	--				
13. Tenure	19.22 (*11.52*)	-.02	.06	.06	-.04	.03	-.02	.00	-.05	-.11**	-.06	-.06	.03	--			
14. Political Beliefs	3.05 (*1.83*)	.00	.00	.13	-.03	-.04	-.06	-.01	-.01	-.04	-.03	-.03	.17***	.06	--		
15. Race	0.72 (*0.45*)	.04	.03	.03	-.03	.03	-.01	-.02	-.03	-.06	-.04	-.01	.03	.16***	-.06	--	
16. Gender	0.53 (*0.50*)	-.07	-.04	.95†	-.09*	-.08†	-.08†	-.08†	-.05	-.07	-.05	.02	.13**	-.01	.03	.01	--
17. Age	41.97 (*10.71*)	-.02	.03	.05	-.02	.02	.00	.07†	-.01	-.09*	-.02	-.03	.04	.86***	.09*	.12**	-.03

N = 557. ***p < .001, **p < .01, *p < .05, $^{\dagger}p$ < .10

Note: Alpha reliabilities are listed on the diagonal. Categorical control variables: Race coded as 0 = Non-white, 1 = White; Gender coded as 0 = Female or other, 1 = Male.

Results

Manipulation checks. We conducted t-tests to determine whether participants understood critical differences in how the algorithms were designed to screen job applicants across the experimental conditions.

We first tested whether participants in the "Fairness through unawareness" condition understood that the algorithms were designed to 'remain blind to applicant demographic characteristics'. Responses to this question were significantly higher in the "Fairness through unawareness" condition ($M = 3.93$, $SD = 1.25$) compared to the Control condition ($M = 2.74$, $SD = 1.43$), $t(269.96) = -7.41$, $p < .001$, indicating that individuals across these groups distinguished between the algorithms' ability to remain blind to job applicant demographic characteristics.

Next, we tested whether participants in the "Demographic parity" condition understood that the algorithms could 'ensure that applicants from different demographic groups would be provided equal outcomes'. Responses to this question did not significantly differ in the "Demographic parity" condition ($M = 3.68$, $SD = 1.22$) compared to the Control condition ($M = 3.45$, $SD = 1.36$), $t(271.15) = -1.46$, $p = .15$, indicating that individuals in these groups did *not* distinguish between the algorithms' ability to ensure equal outcomes for job applicants.[7]

We also tested whether participants in the "Equal opportunity" condition understood that the algorithms were designed to 'ensure qualified applicants from different demographic groups would be provided equal opportunities'. Responses to this question did not significantly differ in the "Equal opportunity" condition ($M = 3.85$, $SD = 1.14$) compared to the Control condition ($M = 3.64$, $SD = 1.25$), $t(272.58) = -1.43$, $p = .16$, indicating that individuals in these groups did *not*

[7] An exploratory t-test revealed that responses to this question were significantly higher ($p < .001$) in the "Demographic parity" condition compared to the "Fairness through unawareness" condition ($M = 3.06$, $SD = 1.48$).

distinguish between the algorithms' ability to ensure equal opportunities for job applicants.[8] To remain thorough in our reporting, we continued to examine all three dummy-coded conditions when performing further analyses.

Fairness Perceptions. To test our first hypothesis (*H1*), we examined the effects of the three dummy-coded conditions on fairness perceptions using a series of multiple linear regression analyses. As shown in Table 3, the effect of "Fairness through unawareness" condition was positive and significant for our three fairness criterion variables ($p < .01$). That is, participants in the "Fairness through unawareness" condition felt the algorithms were higher in procedural, distributive, and overall fairness compared to the Control condition.

In addition, the effect of the "Equal opportunity" condition was negative and significant for overall fairness ($p = .044$), and marginally significant for procedural ($p = .064$) and distributive fairness ($p = .098$). The effect of the "Demographic parity" condition was nonsignificant in all of the regression models.

Overall, these results provide partial support for our hypothesis that people will perceive algorithms as fairer when they are formally designed to mitigate bias (and communicated as such), as compared to an unspecified fairness approach in the Control condition (*H1*). In particular, fairness perceptions were consistently higher in the "Fairness through unawareness" condition compared to the Control condition. However, contrary to expectation, a different pattern of results emerged in the other two experimental conditions. Whereas fairness perceptions did not vary between the "Demographic parity" and Control conditions, they appeared to worsen in the "Equal opportunity" versus Control conditions. Notably, we did not

[8] An exploratory t-test revealed that responses to this question were significantly higher ($p < .01$) in the "Equal opportunity" condition compared to the "Fairness through unawareness" condition ($M = 3.48$, $SD = 1.37$).

observe significant effects for informational fairness for any experimental conditions, possibly hinting at the reduced relevance of this fairness component compared to procedural and distributive fairness.

Table 3. ***Multiple Linear Regression Analyses of Fairness Criterion Variables***

Variables	Procedural Fairness	Informational Fairness	Distributive Fairness	Overall Fairness
Fairness through unawareness	4.01***	1.15	3.32**	3.55**
Demographic parity	-0.79	-0.30	-0.30	-0.35
Equal opportunity	-1.66†	0.08	-1.86†	-2.02*
AI Views	16.67***	11.96***	18.07***	18.55***
Regulation	-0.11	0.18	-1.09	-0.25
Tenure	0.53	1.31	0.17	0.48
Political Beliefs	0.69	-0.01	0.37	-0.35
Gender	-3.13**	-2.61**	-2.86**	-2.84**
Race	-0.60	0.49	-0.01	-0.35
Age	-0.48	-0.66	0.15	0.21
R2	.36	.22	.39	.40

N = 557. Standardized regression coefficients are reported in the table.
***$p < .001$, **$p < .01$, *$p < .05$, $^{\dagger}p < .10$

With respect to the control variables, the effect of AI views was positive and significant across all of the regression models ($p < .001$)—that is, participants who held more positive views toward AI reported more favorable fairness perceptions. We also observed a significant effect for gender for procedural, informational, distributive, and overall fairness (*p* values vary from .002 to .009). To better understand this finding, we split the regression results across the two gender categories, as shown in Table 4. For participants who identified as female, nonbinary, or other, there was a positive and significant effect of "Fairness through unawareness" condition on procedural, distributive, and overall fairness criterion variables ($p < .01$). There was also a

negative and significant effect of "Equal opportunity" condition on distributive (p = .022) and overall fairness (p = .019), and a marginally significant negative effect on procedural fairness (p = .051). By contrast, the effects of the experimental conditions were nonsignificant for participants who identified as male.

Company Perceptions. Next, we tested whether participants' perceptions of the company's attractiveness and trustworthiness varied as a function of experimental conditions. As displayed in Table 5, the effects of the three conditions were nonsignificant. However, the effect of AI views was positive and significant for each criterion measure (p < .001). In addition, there were marginal negative effects for regulation (p = .067) and gender (p = .084) on company attractiveness.

For company trustworthiness, there was a positive and significant effect for political beliefs (p = .014), indicating that liberal-leaning individuals perceived the company as more trustworthy, and a negative and significant effect for gender (p = .009). When we split the regression results across the gender categories (see Table 6), the effect of the "Fairness through unawareness" condition became positive and significant for company attractiveness (p = .01) for participants who identified as women, nonbinary, or other. There was also a negative and significant effect of "Equal opportunity" condition (p = .006) for this gender category. For participants who identified as male, the effects of the experimental conditions were nonsignificant.

Table 4. ***Multiple Linear Regression Analyses of Fairness Criterion Variables by Gender***

	Male[a]				**Female, Nonbinary, or Other**[b]			
Variables	**Procedural Fairness**	**Informational Fairness**	**Distributive Fairness**	**Overall Fairness**	**Procedural Fairness**	**Informational Fairness**	**Distributive Fairness**	**Overall Fairness**
Fairness through unawareness condition	0.09	0.04	0.04	0.07	0.23***	0.07	0.21***	0.21***
Demographic parity	-0.02	0.01	-0.01	-0.03	-0.04	-0.03	-0.01	-0.00
Equal opportunity	-0.01	0.00	-0.01	-0.02	-0.12†	0.01	-0.14*	-0.13*
AI Views	0.61***	0.44***	0.69***	0.67	0.54***	0.47***	0.55***	0.60***
Regulation	0.06	0.07	-0.01	0.03	-0.04	-0.04	-0.06	-0.38
Tenure	0.08	0.14	0.05	0.10	0.00	0.03	-0.03	0.62
Political Beliefs	0.02	-0.05	-0.01	-0.02	0.04	0.03	0.04	-0.03
Race	0.02	0.02	0.03	0.01	-0.06	0.02	-0.04	-0.86
Age	-0.08	-0.09	-0.04	-0.09	0.01	-0.01	0.05	0.13

[a]$n = 259$, [b]$n = 296$. Standardized regression coefficients are reported in the table.
***$p < .001$, **$p < .01$, *$p < .05$, $^{\dagger}p < .10$

Table 5. ***Multiple Linear Regression Analyses of Company Perceptions and Apply Decision***

Variables	Company Attractiveness	Company Trustworthiness	Apply Decision
Fairness through unawareness	0.07	0.06	0.07†
Demographic parity	-0.02	-0.01	0.02
Equal opportunity	-0.05	-0.04	-.08†
AI Views	0.64***	0.67***	0.61***
Regulation	-1.83†	-0.05	-0.02
Tenure	-0.05	-0.02	-0.06
Political Beliefs	0.02	0.08*	0.01
Gender	-0.06†	-0.08**	-0.06
Race	-0.02	-0.04	-0.02
Age	0.06	-0.05	0.05
R2	.41	.46	.37

N = 557. Standardized regression coefficients are reported in the table.
***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Table 6. ***Multiple Linear Regression Analyses of Company Perception Variables by Gender***

	Male[a]		Female, Nonbinary, or Other[b]	
Variables	**Company Attractiveness**	**Company Trustworthiness**	**Company Attractiveness**	**Company Trustworthiness**
Fairness through unawareness	-0.02	0.03	0.14*	0.08
Demographic parity	-0.07	0.02	0.03	-0.04
Equal opportunity	0.06	-0.02	-0.15**	-0.07
AI Views	0.67***	0.70***	0.61***	0.63***
Regulation	-0.08	-0.03	-0.03	-0.05
Tenure	-0.01	0.06	-0.09	-0.10
Political Beliefs	0.04	0.07	0.02	0.09†
Race	0.03	-0.02	-0.07	-0.05
Age	0.02	-0.12	0.10	0.02

[a]*n* = 259, [b]*n* = 296. Standardized regression coefficients are reported in the table.
***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Apply Decision. As shown in Table 5, we also tested the effect of the conditions on participants' likelihood of applying for the open job position at the company. The effect of "Fairness through unawareness" condition was positive and marginally significant (p = .075). In contrast, there was a negative and marginally significant effect of the "Equal opportunity" condition. Together, these results suggest that participants in the former condition were more likely to apply for the job compared to participants in the Control condition. In addition, the effect of AI views was positive and significant (p < .001).

Mediation testing. We predicted that procedural fairness perceptions would mediate the effect of experimental condition on attitudes toward the company, including the decision to apply for the job (*H2-3*). To test these predictions, we conducted parallel mediation analyses (SPSS Model 6 – 4 mediators, controls included; 5000 iterations; Hayes, 2013). This approach allowed us to simultaneously examine the various fairness measures as mediators (vs. testing each mediator individually using separate regression models).

Fairness through unawareness

First, we tested the fairness mediators on "Fairness through unawareness" and company attractiveness. "Fairness through unawareness" significantly and positively predicted procedural fairness (p = .001) but none of the other fairness variables. When "Fairness through unawareness" and the fairness mediators were included in the same regression model, the effects of distributive, informational, and overall fairness were positive and significant (p values vary from .001 to .034), the effect of procedural fairness was positive and marginally significant (p = .09), and the effect of "Fairness through unawareness" was nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that did not include zero, CI = [0.04, 0.15].

Next, we tested company trustworthiness. When "Fairness through unawareness" and the fairness mediators were included in the same regression model, the effects of procedural, distributive, informational, and overall fairness were positive and significant (p values vary from .001 to .013) and the effect of "Fairness through unawareness" was negative and marginally significant ($p = .066$). A bootstrap test produced a confidence interval for the overall indirect effect that did not include zero, CI = [0.04, 0.14].

For the decision to apply for the job, when "Fairness through unawareness" and fairness mediators were included in the same regression model, only the effects of distributive ($p = .005$) and overall fairness ($p < .001$) were significant (and both were positive). A bootstrap test produced a confidence interval for the overall indirect effect that did not include zero, CI = [0.03, 0.15].

Demographic parity

Second, we tested the fairness mediators on "Demographic parity" and company attractiveness. When "Demographic parity" and the fairness mediators were included in the same regression model, the effects of distributive, informational, and overall fairness were positive and significant (p values vary from .001 to .030), and the effects of procedural fairness and "Demographic parity" were nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that included zero, CI = [-0.04, 0.08].

For company trustworthiness, when "Demographic parity" and the fairness mediators were included in the same regression model, the effects of procedural, distributive, informational, and overall fairness were positive and significant (p values vary from .001 to .011), and the effect of "Demographic parity" was nonsignificant. Again, a bootstrap test

produced a confidence interval for the overall indirect effect that included zero, CI = [-0.04, 0.06].

For the decision to apply for the job, when "Demographic parity" and the fairness mediators were included in the same regression model, the effects of distributive ($p = .005$) and overall fairness ($p < .001$) were positive and significant, and the remaining variables were nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that included zero, CI = [-0.04, 0.08].

Equal opportunity

Finally, we tested the fairness mediators on "Equal opportunity" and company attractiveness. When "Equal opportunity" and the fairness mediators were included in the same regression model, the effects of distributive, informational, and overall fairness were positive and significant (p values vary from .001 to .028), and the effects of procedural fairness and "Equal opportunity" were nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that included zero, CI = [-0.07, 0.05].

For company trustworthiness, when "Equal opportunity" and the fairness mediators were included in the same regression model, the effects of procedural, distributive, informational, and overall fairness were positive and significant ($p = .001$), and the effect of "Equal opportunity" was nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that included zero, CI = [-0.06, 0.05].

For the decision to apply for the job, when "Equal opportunity" and the fairness mediators were included in the same regression model, the effects of distributive ($p = .005$) and overall fairness ($p < .001$) were positive and significant, and the remaining variables were

nonsignificant. A bootstrap test produced a confidence interval for the overall indirect effect that included zero, CI = [-0.07, 0.05].

In summary, the mediation results provide partial support for our predictions by demonstrating that "Fairness through unawareness" increased perceptions of procedural, distributive, informational, and overall fairness. In turn, these fairness perceptions enhanced respondents' attitudes toward the company and likelihood of applying for the job (*H2-3*). However, contrary to expectation, we did not observe significant mediation effects for the "Demographic parity" or "Equal opportunity" conditions, raising doubts about the suitability of these two algorithmic criteria in the hiring context.

Conclusions and Future Work

The fields of computer science, information systems, and management all regard algorithmic fairness as being relevant to user adoption of new technologies such as AI-enabled decision-making tools. However, there is a gap between how researchers devise fairness criteria and their understanding by the general public. In the present study, we find consistent evidence that prospective job applicants consider "Fairness through unawareness" to be superior to the lack of any fairness criteria whatsoever. That is, people perceive hiring algorithms with a blind approach to mitigating bias as higher in procedural (and distributive) fairness compared to no explicit fairness approach. Perhaps most significantly, we observed a mediation pattern for "Fairness through unawareness" suggesting that this approach has the highest potential to attract people toward a company.

Additional findings suggest that women and those who identify outside of the gender binary may be more supportive of "Fairness through unawareness" than men, and view companies who advertise this fairness criterion in job postings positively. Further, having

positive general views toward AI appears to enhance perceptions toward AI-enabled hiring algorithms irrespective of which fairness criteria are applied. One implication of this finding is that people who do not have pre-established views toward AI or people who possess negative views about AI may be less responsive to algorithmic fairness criteria.

Contrary to our predictions, we do not find support that individuals can distinguish well between more advanced fairness criteria like "Demographic parity" and "Equal opportunity" as compared to the Control condition. Because the manipulation checks for these two conditions were nonsignificant, the validity of the observed findings is called into question. With this concern in mind, it is interesting that we observed different results across the "Demographic parity" and "Equal opportunity" conditions. While the results for "Demographic parity" were consistently nonsignificant, we uncovered the surprising finding that individuals in the "Equal opportunity" condition perceived the algorithms to be relatively lower in procedural, distributive, and overall fairness compared to the Control condition—suggesting that algorithms designed to ensure equal opportunities may have a detrimental impact in the hiring context. More testing is needed to understand this finding, but this seems to indicate a potential gap between the ever-growing list of fairness criteria generated by researchers and their understanding by everyday people interacting with algorithms.

For future work, it would be useful to study algorithms in more applied hiring settings, such as a job application platform (e.g., Indeed, LinkedIn). For future lab experiments, it could be beneficial to include a human condition to serve as the reference group in order to determine whether there are meaningful differences between how people perceive algorithmic fairness criteria and manual human screening. Based on previous empirical findings (Jussupow et al. 2020; Logg et al. 2019), we expect that people will prefer humans over algorithms to make

decisions despite the positive effects we observed with the “Fairness through unawareness” criterion. Lastly, future research could also explore whether individual perceptions of fairness criteria vary across different occupational and societal contexts, such as hiring versus healthcare, criminal justice, or lending. Doing so would help to determine whether contextual characteristics moderate how individuals perceive and respond to algorithmic fairness criteria.

References

Araujo, T., de Vreese, C., Helberger, N., Kruikemeier, S., van Weert, J., Bol, N., Oberski, D., Pechenizkiy, M., Schaap, G., & Taylor, L. (2018). Automated decision-making fairness in an AI-driven world: Public perceptions, hopes and concerns. *Digital Communication Methods Lab*. http://www.digicomlab.eu/reports/2018_adm_by_ai/

Awwad, Y., Fletcher, R., Frey, D., Gandhi, A., Najafian, M., & Teodorescu, M. (2020). Exploring fairness in machine learning for international development. *CITE MIT D-Lab*.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. *Proceedings of the Conference on Human Factors in Computing Systems*, 1-14. https://doi.org/10.1145/3173574.3173951

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386-400. https://doi.org/10.1037/0021-9010.86.3.386

Colquitt, J. A. (2012). Organizational justice. In S. W. J. Kozlowski (Ed.), *Oxford Handbook of Organizational Psychology* (pp. 526-547). Oxford University Press.

Colquitt, J. A., & Rodell, J. B. (2011). Justice, Trust, and Trustworthiness: A Longitudinal Analysis Integrating Three Theoretical Perspectives. *Academy of Management Journal*, *54*(6), 1183–1206. https://doi.org/10.5465/amj.2007.0572

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable Artificial Intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(1), e1391.

Conlon, D. E., Porter, C. O. L. H., & Parks, J. M. (2004). The fairness of decision rules. *Journal of Management*, *30*(3), 329-349. https://doi.org/10.1016/j.jm.2003.04.001

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

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263.

Grgić-Hlača, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. *Proceedings of the 2018 World Wide Web Conference*, 903-912. https://doi.org/10.1145/3178876.3186138

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems,* 3323-3331. https://dl.acm.org/doi/10.5555/3157382.3157469

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring Attraction to Organizations. *Educational and Psychological Measurement*, *63*(6), 986-1001. https://doi.org/10.1177/0013164403258403

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. (2020). *Proceedings of the 28th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2020_rp/168

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial Disparities in Automated Speech Recognition. *Proceedings of the National Academy of Sciences*, *117*(14): 7684-7689. https://doi.org/10.1073/pnas.1915768117

Kokkodis, M., Papadimitriou, P., & Ipeirotis, P. G. (2015). Hiring behavior models for online labor markets. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 223-232.

Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. *Advances in Neural Information Processing Systems*, 30.

Langer, M., König, C. J., Sanchez, D. R. P., & Samadi, S. (2020). Highly automated interviews: Applicant reactions and the organizational context. *Journal of Managerial Psychology*, *35*(4), 301-314.

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' Fairness Perceptions of Algorithm-Driven Hiring Procedures. *Journal of Business Ethics*, *188*, 125-150. https://doi.org/10.1007/s10551-022-05320-w

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*. *5*(1). https://doi.org/10.1177/2053951718756684

Lee, M. K., Kim, J. T., & Lizarondo, L. (2017). A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3365-3376. https://doi.org/10.1145/3025453.3025884

Leventhal G. S. (1980) What Should Be Done with Equity Theory? In Gergen K. J., Greenberg, M. S., & Willis R. H. (Eds.), *Social Exchange* (pp. 27-55). Springer.

Logg, J. M., Minson, J.A., & Moore, D.A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

McKinsey Global Institute. (2023, June 14). The economic potential of generative AI: The next productivity frontier. Retrieved from https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20economic%20potential%20of%20generative%20ai%20the%20next%20productivity%20frontier/the-economic-potential-of-generative-ai-the-next-productivity-frontier.pdf

McKnight, D. H., Choudhury, V., & Kacmar, C. J. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *Journal of Strategic Information Systems, 11*, 297-323.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35.

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics*, *181*, 1083-1095. https://doi.org/10.1007/s10551-021-04939-5

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149-167.

Paul, A., & Ahmed, S. (2023). Computed compatibility: examining user perceptions of AI and matchmaking algorithms. *Behaviour & Information Technology*, *43*(5), 1002-1015. https://doi.org/10.1080/0144929X.2023.2196579

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: a review, critique, and design agenda. *Human–Computer Interaction*, *35*(5-6), 1-31

Rudman, W., Hart-Hester, S., Richey, J., & Jackson, K. (2016). Hiring for competency: Hiring to not fail vs. hiring to succeed. *Perspectives in Health Information Management*, 1-6. https://www.proquest.com/scholarly-journals/hiring-competency-not-fail-vs-succeed/docview/1810274891/se-2

Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2019). How do fairness definitions fare?: Examining public attitudes towards algorithmic definitions of fairness. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 99-106. https://doi.org/10.1145/3306618.3314248

Scurich, N., & Krauss, D. A. (2020). Public's views of risk assessment algorithms and pretrial decision making. *Psychology, Public Policy, and Law*, *26*(1), 1-9. http://dx.doi.org/10.1037/law0000219

Smith, A. (2018). Public attitudes toward computer algorithms. *Pew Research Center*. https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/

Tarafdar, M., Teodorescu, M., Tanriverdi, H., Robert, L., & Morse, L. (2020). Seeking ethical use of AI algorithms: Challenges and mitigations. *Proceedings of the Forty-First International Conference on Information Systems (ICIS)*, India.

Teodorescu, M.H., Makridis, C. (2024, February 15). Fairness in machine learning: Regulation or standards? *Brookings Institution*, *Center on Regulation and Markets*. https://www.brookings.edu/articles/fairness-in-machine-learning-regulation-or-standards

Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human-ml augmentation. *MIS Quarterly*, *45*(3b), 1483-1499.

Teodorescu, M. H., Ordabayeva, N., Kokkodis, M., Unnam, A., & Aggarwal, V. (2022). Determining systematic differences in human graders for machine learning-based automated hiring. *Brookings Working Paper Series*. https://www.brookings.edu/wp-content/uploads/2022/06/Determining-systematic-differences-in-human-graders-for-machine-learning-based-automated-hiring.pdf

Teodorescu, M. H., & Yao., X. (2021). Machine Learning Fairness is Computationally Difficult and Algorithmically Unsatisfactorily Solved. *IEEE High Performance Extreme Computing Conference (HPEC)*, 1-8, Waltham, MA.

van Berkel, N., Goncalves, J., Hettiachchi, D., Wijenayake, S., Kelly, R.M., & Kostakos, V. (2019). Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study. *Proceedings of the ACM on Human-Computer Interaction*, 3, Article 28. https://doi.org/10.1145/3359130

van den Bos, K., Lind, E. A., & Wilke, H. A. M. (2001). The psychology of procedural and distributive justice viewed from the perspective of fairness heuristic theory. In R. Cropanzano (Ed.), *Series in applied psychology. Justice in the workplace: From theory to practice* (p. 49-66). Lawrence Erlbaum Associates Publishers.

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 205395171774353. https://doi.org/10.1177/2053951717743530

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. *Proceedings of the CHI Conference on Human Factors in Computing*, 1-14. https://doi.org/10.1145/3313831.3376813

Wrzesniewski, A., McCauley, C., Rozin, P., & Schwartz, B. (1997). Jobs, careers, and callings: People's relations to their work. *Journal of Research in Personality*, *31*(1), 21-33.

The Center on Regulation and Markets at Brookings provides independent, non-partisan research on regulatory policy, applied broadly across microeconomic fields. It creates and promotes independent economic scholarship to inform regulatory policymaking, the regulatory process, and the efficient and equitable functioning of economic markets.

Questions about the research? Email communications@brookings.edu.
Be sure to include the title of this paper in your inquiry.