Artificial Intelligence Review
https://doi.org/10.1007/s10462-022-10193-6

People versus machines: introducing the HIRE framework

Paris Will[1,2] · **Dario Krpan**[1] · **Grace Lordan**[1]

Accepted: 13 April 2022

Abstract
The use of Artificial Intelligence (AI) in the recruitment process is becoming a more common method for organisations to hire new employees. Despite this, there is little consensus on whether AI should have widespread use in the hiring process, and in which contexts. In order to bring more clarity to research findings, we propose the HIRE (Human, (Artificial) Intelligence, Recruitment, Evaluation) framework with the primary aim of evaluating studies which investigate how Artificial Intelligence can be integrated into the recruitment process with respect to gauging whether AI is an adequate, better, or worse substitute for human recruiters. We illustrate the simplicity of this framework by conducting a systematic literature review on the empirical studies assessing AI in the recruitment process, with 22 final papers included. The review shows that AI is equal to or better than human recruiters when it comes to efficiency and performance. We also find that AI is mostly better than humans in improving diversity. Finally, we demonstrate that there is a perception among candidates and recruiters that AI is worse than humans. Overall, we conclude based on the evidence, that AI is equal to or better to humans when utilised in the hiring process, however, humans hold a belief of their own superiority. Our aim is that future authors adopt the HIRE framework when conducting research in this area to allow for easier comparability, and ideally place the HIRE framework outcome of AI being better, equal, worse, or unclear in the abstract.

Keywords Artificial intelligence · Recruitment · Hiring · Diversity

1 Introduction

In the workplace, hiring decisions are constantly being made. Traditionally, humans have been the main agents in these critical decisions. Recently, however, there has been a shift towards the use of technology—in particular, Artificial Intelligence (AI)—to assist in workplace decision making processes (Nica et al. 2019), including recruitment. The shift

✉ Paris Will
p.s.will@lse.ac.uk

[1] Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

[2] London School of Economics, 4.01/4.03 Connaught House, 65 Aldwych, London WC2B 4DS, UK

has been rapid, with 37% of businesses adopting AI, a figure that increased 270% from 2015 to 2019 (Gartner 2019).

With respect to recruitment, AI has been a disruptor; 63% of talent acquisition professionals indicate that AI has changed the recruitment process in their organisation (Korn Ferry 2018), and 34% of recruiters believe that AI will be extremely important in shaping the future of hiring practices (Spar et al. 2018). In addition, there is a growing literature in management that is evaluating AI with consideration to how it impacts the recruitment process. We acknowledge an earlier review on this topic by Kochling and Wehner (2020), which highlights the literature debate regarding whether algorithms will reduce or amplify bias in hiring decisions. We expand upon this research by proposing an approach to assess whether AI should be used in the recruitment process. Specifically, we propose the HIRE framework. This new framework can be used to evaluate whether AI is better, equally good, or worse than humans in the hiring context. We envisage that this framework can be used in the literature evaluating AI and recruitment going forward to allow for easier comparison across studies. The crux of the HIRE framework is that it puts emphasis on whether AI is making equally good or better decisions than human recruiters. We note that this salience is important, given the flurry of media attention describing AI as biased (Parikh 2021; Holmes 2019; Lohr 2021) in a manner that would suggest that AI in hiring is a worse alternative to the status quo of human hiring.

In the present article, we demonstrate the HIRE framework by conducting a systematic literature review on the empirical evidence regarding the suitability of AI for hiring purposes. We use the HIRE framework in conjunction with the PRISMA (Moher et al. 2009) system for conducting and reporting systematic reviews. While PRISMA offers protocols and guidelines to complete a summary of research, it does not give specific guidance on the interpretation of research results. As such, for this topic, the HIRE framework is needed to make evaluations regarding the suitability of AI hiring through comparison to human methods of hiring. We next overview the HIRE framework and then present the systematic literature review that compares humans and AI in the context of this framework.

2 Organising framework

We propose a novel framework for assessing the usage of Artificial Intelligence in the hiring process, called HIRE (Humans, (Artificial) Intelligence, Recruitment, Evaluation). This framework describes the decision-making agents that will be compared, the context it can be used in, and the way it will be evaluated.

"Humans" are used as the comparator to judge outcomes of AI against in the context of this framework, given they represent the current status quo. In order to assess whether we should adopt AI over human decisions in the recruitment process, we evaluate the outcomes of AI hiring comparatively to those of human hiring methods.

"Artificial Intelligence" in HIRE refers to any algorithm which has been trained to make automated hiring decisions (Haenlein and Kaplan 2019; Kallem 2012). Such algorithms define the set of rules used to transform data input into decision output. In hiring algorithms, the model is trained on pre- and post- hire data from previous applicants to make predictions about the hireability of future applicants (Kuncel et al. 2014). More specifically, the algorithm learns which factors from the previous applicants line up with positive outcomes such as job performance, and subsequently can make future predictions regarding

which applicants will be the best hires. These algorithms can be trained to simulate human hiring decisions or improve performance and diversity outcomes of the hired applicants.

"Recruitment" in HIRE specifies that we are interested in assessing AI usage over the entire recruitment process ranging from how candidates are ranked on a search engine, to interview or CV evaluation, and including any other aspect of the recruitment process up until the hiring decision is made.

"Evaluation" refers to the procedure that compares human and AI recruitment using the categories of better, equal, worse, or unclear. *Better* means that the outcomes of AI are more preferable than the human outcomes. In this context, better may refer to improving efficiency within the hiring process (Upadhyay and Khandelwal 2018), sourcing the best talent (Geetha and Bhanu 2018) or reducing biases to promote workplace diversity (Sanchez-Monedero et al. 2020; Raghaven et al. 2020). *Equal* means that the outcomes of AI in the hiring context are not significantly different from human outcomes, whereas *worse* means that the outcomes of AI are less preferable than human outcomes. Finally, *unclear* means that the data is inconclusive or that AI and human outcomes were not directly compared. For AI to be better than humans, it is also contingent on AI being deemed ethical. As such, we also specifically focus on the differences in perceptions of AI with respect to ethics, in addition to presenting the evidence on the effectiveness of AI with respect to outcomes of the hiring process.

3 Review methodology

In conducting this systematic review, we followed PRISMA guidelines. We report our results in in line with the proposed HIRE framework. This allows us to assess each empirical finding within the topic of Artificial Intelligence hiring and put emphasis on whether AI was deemed better, worse, or equal to human hiring.

3.1 Eligibility criteria

All empirical studies which assess an aspect of algorithmic hiring written in the English language were included. Experiments were limited to participants who are part of the adult human population (18 years or older). All papers published between 2005 and 2021 are included. 2005 was chosen as a commencement date due to the identified emergence of Artificial Intelligence in the workplace during this time (Ganguly et al. 2005).

3.2 Literature search

The electronic database PsycINFO was searched using keywords relating to the topics of Artificial Intelligence, hiring, and inclusion. PsycINFO was chosen as it is a leading source for information in psychological sciences and has broad coverage on the topic of Artificial Intelligence (APA PsycInfo 2022). We developed the keywords through discussions where we identified synonyms and different spellings of words linked to each topic and refined them to achieve terms general enough to gather the most relevant papers. The keywords were searched for in the title or abstract of the papers and are shown below in Table 1. The database search was repeated prior to submission to ensure the most recent papers were also included. Manual search strategies were also employed to ensure all relevant papers were identified and included. These manual strategies included searching through reference lists and papers citing

Table 1 Search terms used in PsychINFO database search

Search string
TITLE/ABSTRACT: (artificial* intelligen* OR AI OR tech* OR data-driven OR machine learn* OR digital OR algorithm* OR neural network OR robot* OR code* OR compute*) AND (hire* OR hiring OR recruit* OR employ* OR candidate OR resume OR curriculum vitae OR CV OR talent management OR interview* OR job applica* OR job eligibility OR job screen* OR advert* OR job post* OR job list* OR job specification OR job description OR job requirement OR job role* OR job responsibilit* OR job function OR headhunt* OR talent search OR job suitab* OR job competen*) AND (inclusion OR inclusive OR include OR divers* OR divergent OR bias* OR fair* OR equal* OR gender OR race OR racial OR stereotyp* OR exclusi* OR exclude OR heterogeneity OR homogeneity OR disparity OR ethic* OR demographic OR age)

the relevant included papers. Finally, Google Scholar was used to check for any additional citations in line with the eligibility criteria.

3.3 Study records

The keyword search results were saved to one of the author's Ovid account for the duration of the screening process. During the selection process, all titles and abstracts were reviewed. Papers that appeared to fit the eligibility criteria were saved to another folder, and pdfs were obtained for further screening. Once eligibility fit was determined, data regarding methodology and results were extracted from each paper. We extracted information on the population, number of participants, task, conditions, measures, statistical analysis employed, and key findings with significance level and effect size if reported or could be calculated. In the absence of effect sizes, we report percentage differences.

3.4 Paper themes and data analysis

Once information was extracted from all relevant studies, the papers were sorted into themes. Themes were identified by grouping papers together on the basis of the outcome being assessed. We report results in line with our proposed framework; AI hiring is compared to human hiring and deemed as better, equal, worse, or unclear. We report significant results at the $p < 0.05$ level. Where relevant, we discuss effect sizes in line with Cohen's (1962, 1988) classifications shown in Table S1 of the Supplementary Information.

4 Results

In this section we describe the results of this systematic review, starting with the themes identified through the search. We conclude by applying the HIRE framework to each paper, and stating clearly whether AI is deemed better, worse or equal to humans in the paper's specific context. To ease the summary of the papers, we group the findings of each paper into themes based on their outcomes.

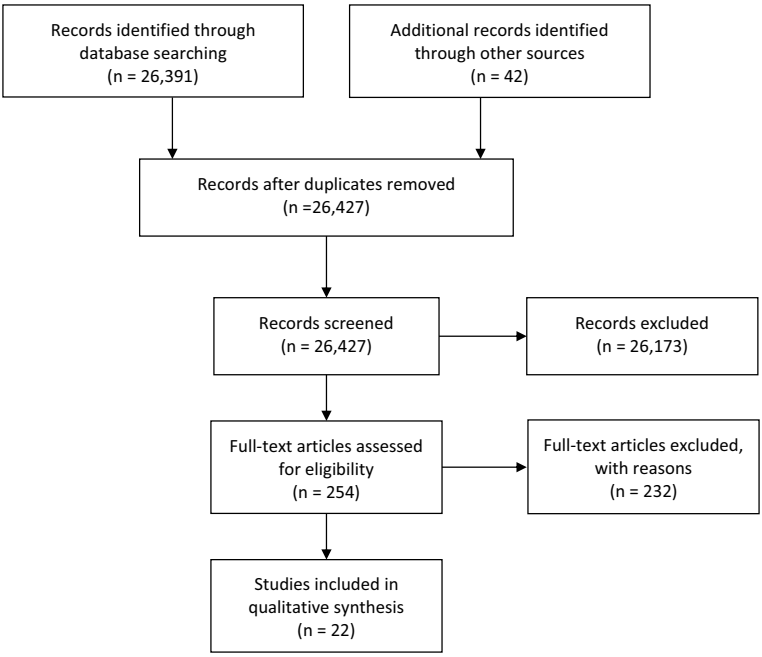

Fig. 1 PRISMA flow diagram illustrating the records identified. Reasons for exclusion were topic not being specific to hiring/artificial intelligence, or non-quantitative empirical papers

4.1 Themes identified

A total of 22 quantitative empirical studies adhered to our inclusion criteria (see Fig. 1) The four themes that emerged from papers groupings are; efficiency, performance, diversity, and perceptions. Sub-themes emerged in the perception section consisting of ethical perceptions, organisational perceptions, perceptions of use, emotional perceptions, and additional perceptions. These themes accounted for all papers, with some papers falling into multiple themes. We evaluate findings of each theme in line with the HIRE framework.

For the efficiency section, we look at how AI can be used to simulate human hiring decisions. We define efficiency as the ability to reach an outcome using the least amount of resources possible. For this topic, the resources used include time and cost of hiring. If AI can make more efficient decisions which are stated to be lower in time and/or cost, it is deemed better. If AI can simulate human hiring decisions, it is considered equal to human hiring. If AI is inaccurate in simulating human hiring, it is worse. We note that this is a cautious approach to take with making conclusions on the efficiency of AI, as AI simulating human hiring decisions in theory will be more efficient due to the rapid nature of algorithmic decision making compared to human decision making.

For the performance section, we evaluate the performance outcomes of candidates hired through AI or human methods. If AI results in better performance outcomes than human hiring, it is deemed as better. If the performance outcomes of candidates hired through AI or human methods are statistically indistinguishable, it is equal. If performance outcomes are worse in AI hiring, it is evaluated as worse.

In the diversity theme, we look at how diversity outcomes differ between AI and human hiring. AI is considered to be better if this type of hiring results in a more diverse group of hired employees. AI is considered equal to human hiring if diversity outcomes are statistically indistinguishable. Lastly, AI is considered worse if it promotes selection of less diverse hires.

For the perception sub-themes, we evaluate how people view AI hiring. AI is considered better if people have more positive perceptions of AI than human hiring. If perceptions towards AI and human hiring are statistically indistinguishable, it is considered to be equal. Finally, if perceptions towards AI hiring is more negative, it is considered worse.

In all of the above themes, findings can also be considered unclear. This can occur when AI is not compared directly to a human hiring method, or the statistical findings are inconclusive or unable to be replicated.

4.2 Efficiency of AI hiring

The potential for AI to automate the hiring process and produce cost-savings is contingent on the ability of AI to replicate human decisions in hiring. Four papers assess the efficiency of algorithmic hiring by whether AI can simulate human hiring decisions (Naim et al. 2018; Stein 2018; Bergman et al. 2020; Horton, 2017). An overview of the studies is shown in Table 2.

Naim et al. (2018) finds that trained algorithms can predict human hiring decisions to a large extent. The authors developed and assessed machine learning algorithms to use verbal and nonverbal behaviours in job interviews in order to predict human interview scores and human rated interview-specific traits. They trained support vector regression (SVR) and least absolute shrinkage and selection operator (LASSO) regression algorithms on the prosodic, lexical, and facial features extracted from audio-visual videos of university students seeking internships. An additional group of participants rated the interview videos in terms of interview performance and additional traits on a 7-point Likert scale. Prediction accuracy was measured by correlating the human ratings to the predicted ratings from the algorithm and estimating AUC (area under the curve) which assesses how well the model can separate true positives from true negatives, as a proxy for how correct predictions are. The results indicated that the algorithms could predict overall interview score ($r = 0.62$, $AUC > 0.76$) and whether the applicant would be recommended for hiring (r ranged between 0.64–0.65, $AUC > 0.78$). Additionally, the model was able to predict some interview-specific traits, such as excitement (r ranged between 0.75–0.79, $AUC > 0.88$), engagement (r ranged between 0.74–0.75, $AUC > 0.84$), and friendliness (r ranged between 0.70–0.73, $AUC > 0.80$). Both the correlation and AUC values from these predictions are large in effect size. However, the algorithms did not perform as well in predicting other traits such as calmness (r ranged between 0.30–0.38, $AUC > 0.60$), level of stress (r ranged between 0.26, $AUC > 0.57$), and eye-contact (r ranged between 0.26–0.33, $AUC > 0.62$), which indicate small-medium effect sizes.

Stein (2018) shows that algorithms modelling cultural compatibility can make predictions which are largely related to hiring outcome, implying that they are good substitutes for humans. The author assesses whether linguistic similarity of a candidate with the hired employees in a firm can predict the job applicants' chance of being hired, and whether this similarity can be algorithmically modelled. Linguistic similarity is a measure of cultural compatibility and is measured through the frequencies of words people use. Among applicants and hired employees in a mid-sized technology firm, data

Table 2 Papers assessing efficiency of AI hiring

Citation	Key finding (s)	HIRE evaluation
Naim et al. (2018)	In comparing algorithmic ratings to human ratings of job candidates, both types of ratings were highly correlated in terms of overall interview score, candidate recommendation, and assessment of candidate traits such as excitement, engagement, and friendliness	Equal
Stein (2018)	Algorithm found to be able to predict human hiring outcomes based on linguistic similarity of candidates with current employees	Equal
Bergman, Li, & Raymond, (2020)	Hiring algorithm can moderately predict human hiring outcome. The algorithmic candidate recommendations found to increase proportion of candidates hired after interview by 20% compared to human recruiter candidate recommendations	Better
Horton, (2017)	The use of algorithmic recommendations found to increase hiring fill rate for technical job openings by 20%	Better

from job application questions was used to model linguistic similarity through logistic regression algorithms. The authors transformed job application responses into a term frequency inverse document (TF-IDF) statistic and then measured similarity of words between applicants and employees. They found that depending on the control measures included, a one standard deviation increase in linguistic similarity increases a candidate's likelihood of being hired by between 33–53%, seen through significant hiring odds ratios of 1.331–1.529, $p < 0.05$.

Bergman et al. (2020) find that algorithms can predict hiring outcomes in a modest way, and that the data-driven method outperforms human recruiter predictions. Specifically, they train supervised learning (SL) and upper confidence bound (UCB) algorithms on applicants' demographics, education, and work history to predict whether they will be hired by a human after an interview. The data comes from a Fortune 500 company which had been interviewing candidates based on recruiter's recommendations. In this company, 10% of applicants are hired after being interviewed, thus, the algorithmic approach was studied to see if they could increase this percentage and improve efficiency through interviewing fewer candidates. The algorithms produced a score for the likelihood of an applicant being hired and when correlated to actual hiring outcome yielded small positive and significant correlations (r ranged between 0.214-0.233, all $p < 0.001$). Additionally, using algorithmic recommendations for the sample of candidates interviewed increased the hiring rate to 30%. This means that the company could use these algorithms over recruiter recommendations to find a suitable hire with the added bonus of conducting 20% fewer interviews, resulting in time savings.

Finally, Horton (2017) assessed how the introduction of algorithmic candidate recommendations on an online labour market impacts candidate hiring outcomes. The algorithmic recommendations were computed by measuring a candidate's relevance and ability. Relevance was calculated by the degree of overlap of skills listed on a candidate's page and the skills required for job. Ability included the worker's test score, feedback ratings, and past earnings. The actual algorithm was "black box" so the exact type of computational process is unknown. Employers were randomly assigned to the treatment group where they were given algorithmic candidate recommendations for their job postings, or the control group where they were not given any recommendation. Among the employers who were given algorithmic recommendations, hiring rate increased by 20%. However, this pattern only emerged for technical jobs, and algorithmic recommendations did not increase non-technical job fill rates. The authors argue that this is likely because technical job openings typically have fewer applicants and are less price sensitive than non-technical job openings. The algorithm helped technical jobs to have higher numbers of suitable applicants, promoting hiring for those roles. The algorithm also screens for past earnings as an aspect of candidate ability, so the algorithmically recommended candidates are typically higher cost, and because the technical jobs are less price-sensitive, they were more likely to hire these more expensive algorithmically recommended candidates. Thus, the use of AI hiring in this situation helped to increase efficiency for technical jobs through increasing the fill rate of positions.

The results of these four studies indicate that hiring algorithms can be designed to produce outputs which are equal to or better than human hiring. The ability of the algorithms to predict hiring outcomes varied in size, but even at the smallest effect size in Bergman, Li, & Raymond (2020), AI still outperformed human prediction methods. From these results, it would seem that AI can be used to model or improve human hiring outcomes, thus producing time cost-savings in the recruitment process.

Table 3 Papers assessing the performance outcomes of AI hiring

Citation	Key finding (s)	HIRE evaluation
Sajjadiani et al. (2019)	Algorithm developed to maximize selecting hires based on performance outcomes overlapped with human recommended hires by 11–29%	Better
Bergman, Li, & Raymond (2020)	Algorithmic candidate recommendation score found to be significantly positively related with whether a new hire receives a promotion. Human recruiter candidate recommendation score was unrelated to whether a new hire receives a promotion	Better

4.3 Performance outcomes of AI hiring

In order to consider if AI can make better hiring decisions in terms of the employee's performance outcomes, two studies assess novel algorithmic methods in predicting workplace talent from an applicant pool (Sajjadiani et al. 2019; Bergman et al. 2020). Refer to Table 3 for an overview of these studies.

Sajjadiani et al. (2019) finds that algorithms can maximise selecting hires on the basis of performance outcomes. They develop a machine-learning hiring technique based on prior data from teaching position applications in a school district. The machine learning method utilised a naïve Bayes classifier to predict work outcomes from data on the applicant's demographics, previous work experience, tenure history, and turnover reasons.

Work outcomes from the hired applicants included turnover and the performance variables of student evaluations, expert observations, and value-added measured by students' scores on standardized tests. The model was trained on 90% of the data and evaluated on the remaining 10%. The model was evaluated by looking at how well the input variables predicted work outcomes, and Heckman regression indicated a number of significant coefficients, indicating that input variables could predict work outcomes. The model was compared to human selection methods by developing a list of recommended hires based on predicted performance and turnover outcomes, and comparing these to the candidates who were actually hired. Depending on the outcome the model was maximized to predict, results showed that the overlap between algorithm recommended hires and actual hires was between 11–29%. Since the algorithm maximized predicted performance outcomes, this suggests that the algorithmic hiring method differs to human hiring in that it may result in hires with higher performance.

Bergman et al. (2020) find that the computed candidate score from hiring algorithms is modestly and positively related to promotion and performance outcome, and that it was an improvement over human methods. They correlated measures of job performance ratings and promotions with the algorithmic and human recommendation score. Performance was rated on a 3-point scale, and promotions followed a binary yes/no outcome. They found that human recruiter recommendations are significantly negatively correlated with job performance rating (r ranged between -0.288 to -0.309, $p < 0.01$), and algorithm recommendations are significantly positively correlated with whether a new hire receives a promotion ($r = 0.132$, $p < 0.01$). This means that candidates selected by human recruiters have worse job performance, and those selected by algorithms are more likely to be promoted,

and these effects are small-medium in size. All other correlations were insignificant due to very small effect sizes, however, the human recommendations were slightly negatively correlated with promotions and the algorithm was slightly positively correlated with performance ratings.

Taken together, these two studies indicate that AI hiring methods are better than human ones in terms of selecting applicants who will have better performance while on the job. Although as shown in Bergman et al. (2020), the ability of AI to predict candidates with positive job performance is limited, but still an improvement over human methods.

4.4 Diversity outcomes of AI hiring

Due to the potential of AI to reduce human cognitive biases in decision-making, it has been examined as a tool to promote diversity and inclusion. Seven studies explore how inclusive algorithmic hiring is, and advance novel approaches to improving inclusivity metrics (Chen et al., 2018; Lambrecht & Tucker, 2019; Allred, 2019; Song, 2018; Sajjadiani et al. 2019; Bergman et al. 2020; Suhr et al. 2020) See Table 4 for an overview of the studies.

Sajjadiani et al. (2019) finds that AI hiring can remove adverse impact for gender and ethnicity. In a Bayesian machine learning approach, they assessed the level of adverse impact[1] in selection decisions. Adverse impact is evaluated by computing whether sensitive variables such as gender and ethnicity predict hiring outcome. While the previous human hiring method showed that gender and ethnicity were in fact predictive of hiring (β female = 0.06, $p < 0.05$; β white = 0.11, $p < 0.01$), the machine learning method yielded a non-significant prediction with coefficients close to zero, implying that no group would be disproportionately hired with the algorithmic method.

On the other hand, two studies (Chen et al., 2018; Suhr et al. 2020) look at how search engine algorithms can cause discrimination in candidate ranking. Chen et al. (2018) find that females are ranked slightly lower than men on search engines. Data was collected from three resume search engines on job titles in 20 U.S. cities. The type of algorithms used in the search engines are unspecified so the authors could only assess the outcomes. On such websites, a higher ranking indicates a better candidate, and recruiters will look at highly ranked applicants first. Although sensitive features were removed from profiles for the rankings, it was found through Mixed Linear Model regressions that female candidates were ranked statistically lower than males ($\beta = -0.042, -0.028,$ & -0.071; $p < 0.05$). The effect was small, however, so in the top 10 rankings, the gender difference was proportionally negligible. The authors also assessed how this effect shows up in specific job contexts. Mann–Whitney U tests revealed that 8.5–13.2% of job titles have significant group gender unfairness, all with men being ranked higher than women (U ranged from 0.01–,0.59, $p < 0.05$). Since the search engine websites use black box algorithms, the reason for this discrimination can only be speculated. The authors hypothesize a hidden feature in the algorithm that is correlated with gender, or that rankings are adjusted based on recruiter clicks and recruiters are biased towards favouring male candidates.

Suhr et al. (2020) find that gender discrimination in candidate selection can be improved by using alternative ranking algorithms. They recruited online participants in a simulated job selection process and had them review ranked candidates and select four in order of

[1] It should be noted that the authors did not intend on reducing adverse impact, and that this was a by-product of the machine learning approach to improving job performance predictions.

Table 4 Papers related to diversity outcomes of AI hiring

Citation	Key finding(s)	HIRE evaluation
Sajjadiani et al. (2019)	Machine learning method of hiring found to have no adverse impact for gender and ethnicity, whereas the human method of hiring was found to have adverse impact for both gender and ethnicity	Better
Allred, (2019)	Algorithm able to reduce racioethnic group differences in general cognitive ability scores compared to previous weighting method	Better
Chen et al. (2018)	On three candidate search engines used by recruiters, female candidates are ranked statistically lower than male candidates	Unclear
Lambrecht and Tucker (2019)	A job advert algorithm designed to be gender neutral was shown to 20% more men than women	Unclear
Song (2018)	Algorithm able to reduce validity shrinkage seen in pareto-optimal methods for AI hiring	Better
Bergman et al. (2020)	Upper-confidence bound algorithm and blind-upper confidence bound algorithm recommends 24% and 14% minority candidates. Human recruiters recommend 9% minority candidates	Better
Bergman et al. (2020)	Static learning algorithm recommends 3% minority candidates, whereas human recruiters recommend 9% minority candidates	Worse
Suhr et al. (2020)	A fair ranking algorithm increased selection of female candidates in comparison to search engine ranking by 2.5–17.35 percentage points	Better

preference to be recommended to a company. Participants were randomly assigned to see candidates which were grouped based on three ranking algorithms and three datasets. In the first ranking algorithm, candidates were ranked by their relevance score from a search engine website; in the second algorithm, candidates were ranked in a random order; and in the third algorithm, candidates were ranked by a fairness ranking algorithm (LinkedIn's *DetGreedy)* which ensures a proportional representation of underrepresented groups. They found that the fair ranking algorithm increased selection of female candidates in comparison to the relevance ranking by the search engine from 2.5–17.35 percentage points depending on the job context and ranked position out of 4, with higher gains for females being selected as the first recommendation. The random ranking algorithm also increased selection of female candidates, but in all cases by less than the fair ranking algorithm. This shows that the way individuals are ranked on hiring search engines can have diversity consequences in terms of actual hiring outcomes.

Lambrecht and Tucker (2019) tested an algorithm for showing job adverts to individuals in STEM which was designed to be gender neutral, however, they find that the advert was displayed to more men than women. The job advert was programmed to be gender neutral by setting the ad targeting settings to both genders. Despite the good intentions behind the algorithm, the job advert was found to be displayed to 20% more men than women ($R^2 = 0.49$, $p < 0.001$). Since an additional analysis showed the ad appealed similarly to both men and women, the authors discovered the effect to be driven by economic factors. The price premium that an advertiser must pay to show ads is more expensive for women than men. This is because women have a higher click to profit ratio, meaning they are more likely to purchase an item after clicking on it. This study emphasises that discriminatory effects can occur even when AI is designed to be fair, and the importance of checking external factors which may impact the outcome of AI. It is important to note, however, that this algorithmic discrimination was not compared to human methods, so it is unclear if gender discrimination outcomes are better or worse than the status quo.

The last three studies propose novel algorithmic methods to improve inclusion in hiring. (Bergman et al. 2020; Allred, 2019; Song, 2018). Bergman et al. (2020) find that depending on the type of algorithm, hiring outcomes for underrepresented individuals can either be greatly improved or moderately worsened. In the static supervised learning (SL) algorithms, the model is trained on a dataset to predict a candidate's likelihood of being hired. In the upper confidence bound (UCB) model, the algorithm values a reduction in uncertainty, and exploration bonuses are given for hiring candidates which have higher standard error due to less reported outcomes. This means that the model favours selecting candidates with less hiring data outcomes on, in order to build up the model to have better predictions for all individuals. Diversity outcomes of the models and the baseline human recruiter method are assessed. In the sample of all applicants, the majority racial groups are White and Asian (79%) and the minority groups are Black and Hispanic (21%). In the human recruiter selection, minority applicants represent 9% of selected candidates. For the SL models, the model decreases minority groups so that they only represent 3% of chosen applicants. However, for the UCB model, the proportion of minority groups increases to be 24% of the selected candidates. In a further extension study of the UCB model, where the model is blinded to race, minority groups drop down to being selected 14% of the time. This means, depending on the type of AI and what data is inputted, it can be much better or slightly worse than humans at selecting underrepresented groups for hire.

Allred (2019) finds that group differences in a cognitive test used for hiring can be substantially reduced through an algorithmic method. The author responds to the problem of racioethnic group differences in the general cognitive ability (GCA) assessment, meaning

that there are differences in test scores based on one's racial group. Despite this, GCA is utilized in many organisations due to the high level of prediction of job outcomes, but this can result in lower selection rates for certain racial groups. To promote fairer selection, Allred (2019) proposes a metaheuristic algorithm to lower group differences in the test scores. Specifically, they use an ant colony optimization algorithm to differentially weight items on the test so that racial group differences are minimized while validity of the measure is maximized. The author used simulated data from archives and meta-analytic estimates of mean differences between groups in the GCA. Compared to the prior approaches to GCA interpretation, the metaheuristic algorithm produced smaller group differences in GCA scores. The effect sizes varied based on level of job complexity,[2] with group differences lowering from a large (d = 0.840-0.845) to small (d = 0.349) effect size in low job complexity, and medium (d = 0.602-0.730) to small (d = 0.442-0.475) in medium and high job complexity. However, the diversity improvements came at the cost of a reduction in cross-validation validity, which was also dependent on job complexity. In low job complexity, validity was 52% lower, in medium job complexity it was 30–31% lower, and in high job complexity it was 25% lower. Thus, at low job complexity where the greatest diversity improvements occur, the greatest reduction in validity also occurs. This is known as the diversity-validity dilemma, and Song (2018) sought to correct it.

Specifically, Song (2018) identify shrinkage as the validity issue in pareto-optimal methods for personnel selection. Shrinkage occurs when a model is over fitted, and so variance explained in a new sample will be smaller than in the original training data. This means that if organisations employ pareto-optimal methods in their personnel selection process, the diversity and validity outcomes may not be as optimal as the model may lose prediction ability in who the best employees are to hire. This study examined how much shrinkage occurs with pareto-optimal methods in personnel selection by assessing the cross-validity through a Monte Carlo simulation with the factors of sample size and job predictors. Job predictors included cognitive ability tests, biodata, conscientiousness, structured interview, and integrity test results. Calibration and validation samples were generated for conditions which varied on sample size and job predictors, then pareto-optimal weights for validity and diversity were calculated for the calibration and validation samples and were plotted against each other to observe amount of shrinkage. The results showed that the validation curve fell beneath the calibration curve, meaning there was substantial shrinkage in validity and diversity. However, even with the shrinkage, diversity was still improved over the status-quo unit weighted method when the sample size was at least 100. In attempt to correct the shrinkage, the author developed an algorithm to achieve regularization where optimization occurs in local predictions and future generalizations. This is done through four steps, where end-points of the pareto-optimal curve are detected where diversity or validity is maximized, then a phi trade-off matrix is built to estimate optimal predictor weights, linear interpolation creates evenly spaced solutions for pareto-optimal points, and finally a sequential least squares programming algorithm is used to find optimal weights for regularization. Through observing a smaller gap in the pareto-optimal curves between calibration and validation samples, they found that diversity shrinkage was smaller for this solution as compared to original pareto-optimal results, showing that validity of such AI can be improved.

[2] Job complexity came from Roth et al.'s (2001) meta-analytic estimates of GCA and job complexity. In their study, job complexity is based on the amount of information processing required for the job.

Among the papers in this section which compare AI to human hiring diversity outcomes, each shows that AI is better at producing more diverse hiring outcomes. The exception comes from the Bergman, Li, & Raymond (2020) paper where AI diversity outcome can be worse than humans when a static supervised learning algorithm is used, or better when a UCB algorithm is used. This points to the importance of using the correct type of AI for maximizing diversity. Two papers in this section do not compare AI to human methods (Chen et al., 2018; Lambrecht & Tucker, 2019) and show that AI recruitment methods can result in poor diversity outcomes. However, because these papers did not compare to status-quo human methods, it is unclear whether these findings would be better or worse than the diversity outcomes in human hiring.

4.5 Perceptions of AI hiring

Twelve of the papers included in this systematic review pertain to perceptions surrounding the use of hiring with AI (Suen et al. 2019; Newman et al. 2020; Lee 2018; Langer et al. 2018, 2019; Kaibel et al. 2019; Kodapanakkal et al. 2020; Oberst et al. 2020; Acikgoz et al. 2020; Noble et al. 2021; Warrenbrand 2021; Bigman 2020). These empirical papers cover a wide range of types of perceptions and will be discussed within the categories of ethical perceptions, organisational perceptions, perceptions of use, emotional perceptions, and additional perceptions.

4.5.1 Ethical perceptions

There are ten papers which study ethical perceptions regarding the use of AI in the hiring process (Suen et al. 2019; Newan et al. 2020; Lee 2018; Langer et al. 2018; Kodapanakkal et al. 2020; Langer et al. 2019; Acikgoz et al. 2020; Noble et al. 2021; Warrenbrand 2021; Bigman 2020). The majority of experiments on this topic assess the perceived fairness of hiring using AI technologies. See Table 5 below for an overview of the studies.

Suen et al. (2019) and Newman et al. (2020) experimentally simulate a job hiring scenario to see how fair the applicants perceive the use of AI in personnel selection. Suen et al. (2019) find that whether an interview is conducted with a human or AI, fairness ratings do not differ. They conducted structured job interviews where participants were assigned to one of three conditions: synchronous video interviews (SVI), asynchronous video interviews (AVI), or asynchronous video interviews using an AI decision maker (AI-AVI). In the SVIs, communication is two-way, meaning the applicants interact with an interviewer over a video call. In contrast, the AVI is a one-way interview where job applicants record interview answers to be evaluated at another time. All applicants completed a questionnaire following the interview where they rated the perceived fairness of the interview process by answering questions adopted from Guchait et al. (2014) on a 5 point Likert scale, and results indicated that there were no significant differences in ratings between the three conditions ($np^2 = 0.004$, $p = 0.482$). While this study does not find evidence that perceived fairness in a hiring context is significantly altered by an algorithmic or human decision maker, Newman et al. (2020) study introduces an additional factor in this relationship and finds that fairness does moderately relate to type of interview.

In Newman et al. (2020) experiment, undergraduate students participated in an asynchronous video interview where they were told that either a human or algorithm would analyse their responses prior to the interview. The authors also manipulated transparency by randomly allocating the participants to low or high transparency. For the

Table 5 Papers assessing the ethical perceptions of AI hiring

Citation	Key finding(s)	HIRE evaluation
Suen et al. (2019)	Applicants had indistinguishable fairness ratings between human and AI hiring methods	Equal
Newan et al. (2020)	In a low transparency condition, algorithmic hiring was perceived as more fair than humans	Better
Newan (2020)	In a high transparency condition, algorithmic hiring was perceived as less fair than humans	Worse
Lee (2018)	Algorithmic hiring decisions found to be viewed as less fair than human hiring decisions	Worse
Acikgoz et al. (2020)	AI interviews rated lower on procedural justice and interactional justice compared to human interviews	Worse
Noble, Foster, & Craig (2021)	AI interviews rated lower on procedural justice and interpersonal justice compared to human interviews	Worse
Warrenbrand (2021)	AI hiring rated lower on distributed and procedural justice than human hiring	Worse
Langer et al. (2018)	No difference in fairness ratings in algorithmic hiring conditions with low or high information given	Unclear
Kodapanakkal et al. (2020)	Data protection found to be strongest driver of moral acceptability of algorithms	Unclear
Langer et al. (2019)	No difference in candidate fairness ratings between AI and human interviews	Equal
Bigman et al. (2020)	Participants are less outraged by hiring discrimination from an algorithm than from a human. Participants attribute higher prejudiced motivation to human than algorithm decision makers. The attribution of bias mediates the effect of discrimination on moral outrage	Better

participants in the high transparency condition, they were shown an additional paragraph about how the human or AI evaluates the interviews. Fairness measurement was adopted from Conlon et al.'s (2004) organisational justice scale, where questions were answered on a 7-point Likert scale. This transparency factor was found to be significant in predicting fairness scores. When participants were given additional information on the interview process, algorithms were found to be less fair than humans to a medium extent ($d = 0.5$, $p = 0.011$). However, when participants were not given additional information about how the human or AI evaluates interviews, this finding was reversed and algorithms were found to have higher ratings of fairness than humans.

While the findings of these studies would be potentially compatible if Suen et al. (2019) used a medium transparency method, as the point where the fairness effect crosses over from low to high transparency and is null, inspection of the methodology for that paper reveals description of the instructions to participants aligns closely with Newman et al. (2020) low transparency condition. Other factors may be involved in these different results, as there were other notable methodological differences. The job questions used in the interview were different, as well as the level of seniority of the job. While Suen et al. (2019) hired for HR managers of a mid-senior level, Newman et al. (2020) recruited undergraduate students for an unspecified future job opportunity. Additionally, the most prominent difference between the methodology of these studies is the time at which applicants filled out the questionnaire on the hiring process. Participants in the Suen et al. (2019) experiment completed the interview and then filled out the questionnaire, whereas applicants in Newman et al. (2020) were briefed on the process and filled out the questionnaire prior to the interview. Thus, it is possible that the experience of completing the actual interview resulted in different fairness ratings.

Another three papers on perceived fairness asked participants to assess a hypothetical hiring scenario (Lee 2018; Langer et al. 2018, 2019). While these experiments do not have real stakes, such as participating in a job interview, they inform what potential applicants might think when evaluating a company, prior to the application stage. Langer et al. (2018) find that there is no significant effect of AI interviews on fairness, regardless of the level of transparency. They manipulated transparency similar to Newman et al. (2020) by providing additional details regarding how the AI program analyses the video and audio information to the participants in the high information condition, and leaving this out for participants in the low information condition. They had participants observe a job interview where the interviewer was a virtual character who interacted with a human applicant, and then asked participants to answer a questionnaire on the process. Fairness was adapted from Warszta (2012) and questions were answered on a 5-point Likert scale. The findings indicated that there was a small but insignificant effect of higher fairness ratings in the high transparency condition ($np^2 = 0.03$, $p > 0.05$).

Langer et al. (2019) also find no difference in fairness ratings between type of interview, but find a moderate effect that level of interview stakes influence fairness ratings. They had participants watch either an automated interview or videoconference and varied stakes by saying the interview was for training and feedback in the low stakes and saying that the interview was real in the high stakes condition. The authors used the same fairness measurement as Langer, Konig, and Fitili (2018) and also found a small but insignificant effect of fairness being lower in automated interviews ($np2 = 0.03$, $p > 0.05$). There was a medium sized significant difference between level of stakes and fairness in that for participants who believed the interview was real, they rated the interview process as less fair ($np^2 = 0.07$, $p < 0.01$).

On the other hand, Lee (2018) finds a large effect that human interview decisions are rated as more fair than AI ones. They recruited participants to evaluate a hypothetical hiring scenario and assessed perceptions around using an algorithm or a human for the initial recruitment stage, such as reviewing resumes and personal statements on a job website. Fairness measurement questions were adopted from Brockner et al. (1994) and Konovsky & Folger (1991), and participants answered questions on a 7-point Likert scale. Participants gave much higher fairness ratings when the decision maker was a human rather than an algorithm ($d = 0.861$, $p < 0.0001$).

Due to the mixed findings for fairness, more recent studies have broken down the meaning of fairness into different components (Acikgoz et al., 2020; Noble et al. 2021; Warrenbrand, 2021). All three studies looked at procedural justice, which in this context is the perceived fairness of the decision making process for hiring, whether through human or AI means. All three studies utilised a vignette style methodology where participants read over a hiring process situation in which a human or AI made the hiring decision. Procedural justice was measured through Bauer et al.'s (2001) 5-point Likert (Acikgoz et al., 2020), Bauer et al.'s (2001) 7-point Likert (Noble et al. 2021), or Colquitt (2001) 5-point Likert scale (Warrenbrand, 2021). Across the three studies, procedural justice was rated moderately to substantially higher in the condition which used a human decision maker for the hiring process (d ranged from 0.27 to 0.80, $p < 0.05$).[3]

Acikgoz et al. (2020) and Noble et al. (2021) also looked at the perceived fairness of how individuals are treated during the hiring process (i.e., interactional or interpersonal justice), measured through Bauer et al.'s (2001) 5-point or 7-point Likert scale. Interactional justice was generally rated as much higher for human decision makers (d ranged from 0.78 to 1.40, $p < 0.001$)[4] (Acikgoz et al. 2020). Interpersonal justice was rated moderately higher for human decision makers (d ranged from 0.26 to 0.64, $p < 0.05$) (Noble et al. 2021). Furthermore, the treatment sub-factor mediated the influence of type of interview on litigation intentions (β=0.09, $p < 0.05$) (Acikgoz et al. 2020). Litigation intentions in this context indicates how likely people were to report chance of seeking legal recourse from the hiring process, and participants who were shown the automated hiring process felt that they were treated worse, thus causing them to be more likely to seek legal recourse.

Warrenbrand (2021) also looked at distributive justice, which in this case refers to applicant's perceptions of the fairness of a hiring decision, comparing their experience to other applicants' experience. Distributive justice was measured on a Colquitt's (2001) 5-point Likert scale. It was found that distributive justice was rated as substantially higher when humans made the hiring decision ($d = 1.13$, $p < 0.001$).

The last three papers discussing ethical perceptions are related to morality and privacy concerns (Kodapanakkal et al. 2020; Langer et al. 2019; Bigman et al. 2020). Langer et al. (2019) find that people have slightly more privacy concern towards AI interviews than human ones. Participants watched either a videoconference or automated interview and then filled out a questionnaire regarding the process. Privacy concern was measured with six items from previous research (Smith et al. 1996; Agarwal et al. 2004; Langer et al. 2018, 2017) using a 7-point Likert scale. The authors found that levels of privacy concern were higher in the automated interview condition ($np^2 = 0.04$, $p < 0.05$), however, this finding is

[3] This is with the exception of the sub-facet of consistency, which was rated higher for AI decision makers in Acikgoz et al. (2020) ($d = .9$, $p < .001$) and Noble et al. (2021) ($d = .36$, $p < .001$).

[4] This is with the exception of the sub-facet of information known, regarding which there was no significant difference between the human and AI conditions ($d = .22$, $p = .10$).

considered to be small in effect size. This may also inform about other ethical aspects of using algorithmic hiring, as Kodapanakkal et al. (2020) extend these findings to show that data protection drives moral acceptability of a hiring algorithm. Moral acceptability was rated by participants on a 0–100 numeric scale. They manipulate outcome favourability of the technology by stating whether the algorithm will increase or decrease the chance of someone finding employment. Data sharing is manipulated by stating whether the data will be shared with no one, a private company, or academic researchers, and data protection is altered by stating whether the data is encrypted and stored securely. They found that data protection was the driving factor in the moral acceptability ($z = 14.68$, $p < 0.001$) of using such algorithms for hiring.

Finally, Bigman et al. (2020) found that people are less morally outraged when an algorithm discriminates than a human. Moral outrage was measured on a 7-point Likert scale using a measure from Sunstein et al. (1998), and attribution of moral outrage was measured through questions created by the authors on a 7-point Likert scale. Across three hiring scenarios where discrimination occurred on the basis of race, age, or gender, algorithms were consistently rated to evoke less moral outrage, with the effect size varying from small to large (d ranged from 0.34 to 0.80, $p = 0.012$). This effect was mediated by the attribution of bias, meaning the perceived motivations behind the discrimination ($b = -0.30$, $p < 0.05$). People attributed substantially higher levels of prejudiced motivation to the human decision than the algorithmic hiring decision ($d = 1.03$, $p < 0.001$).

Overall, the current findings regarding ethical perceptions of AI are mixed, but most findings point towards AI perceptions being equal to human, or worse. Due to this, further research is needed to identify other factors that influence ethical perceptions of AI in a hiring context. From the findings in Langer, Konig & Papathanasiou (2019) and Kodapanakkal et al. (2020) it is possible that concern about data privacy may be a driving factor guiding the ethical perceptions of AI.

4.5.2 Organisational perceptions

How attractive an organisation is perceived to be due to the use of algorithmic hiring is assessed in four studies using hypothetical hiring scenarios (Kaibel et al. 2019; Langer, Konig and Fitili 2018; Langer, Konig, and Papathanasiou 2019; Acikgoz et al. 2020). See Table 6 for an overview of the studies.

Acikgoz et al. (2020) found a moderate effect in that ratings of organisational attraction are lower on automated interviews. In their experiment, participants reviewed a vignette of a job hiring situation and then rated the organisational attractiveness on a 5-point Likert scale from Highhouse et al. (2003). In this study, results showed that participants in the condition where interviews were automated gave moderately lower scores for organisational attractiveness than participants in the condition with human interviewers ($r = 0.29$, $p < 0.001$).

Kaibel et al. (2019) also finds that organisations using AI hiring are considered slightly less attractive than ones using human hiring. They asked participants to evaluate a hiring process where the decision maker was either a human or an algorithm. Organisational attractiveness was measured on Highhouse et al. (2003) scale. Across two studies, they found a small significant effect ($d = 0.375$-0.443, $p < 0.05$) that participants rated organisational attractiveness lower when an algorithm rather than a human made the hiring decision. It was also found that personal uniqueness negatively moderates this relationship; the more an individual considers themselves to be unique, the lower organisational

attractiveness is when an algorithm makes the hiring decision (β coefficient = −0.46, p = 0.002). This suggests that there are individual differences in how applicants view organisations which use algorithmic hiring.

Langer, Konig & Papathanasiou (2019) measured organisational attractiveness on the User Experience Questionnaire (UEQ; Laugwitz et al., 2008) 7-point Likert scale. They also found a small effect of attractiveness being lower on automated interviews (np2 = 0.04, p < 0.05). This effect was mediated by two factors. The first, social presence, meaning the lack of human interaction in automated interviews resulted in lower attractiveness scores. The second, fairness, meaning perceptions of lower fairness in automated interviews also resulted in lower attractiveness ratings. These mediation effects were found to explain 61% of variance in organisational attractiveness scores ($r^2 = 0.61$, p < 0.01).

Langer et al. (2018) extend these findings by including the factor of transparency to test its effect on organisational attractiveness in algorithmic hiring, and find mixed effects. They measure organisational attractiveness on a 5-point Likert scale adapted from Highhouse et al. (2003) and Warszta (2012). These authors found that the amount of information known about AI during the hiring process affects organisational attractiveness in two opposite ways; there is an indirect positive effect of information on organisational attractiveness through the factors of open treatment and information known, but also a negative direct effect of information on organisational attractiveness. This model explained 24% of variance in organisational attractiveness scores. As the authors postulate, these opposing effects might be driven by applicants appreciating the honesty, but being intimidated by the technological aspects of the selection process. It could also be that the amount of information provided was enough to make applicants sceptical, but not enough to explain the methodology so that the participants had no concerns.

It is, however, important to point out that the positive indirect effect that Langer et al. (2018) describe may also be a statistical artefact. The main purpose of a mediation analysis is to explain the psychological mechanism behind the main effect of an intervention on a dependent variable of interest (Yzerbyt et al., 2018)—in this case, the influence of information level on organisational attractiveness. In that regard, obtaining a mediated (i.e., indirect) effect that is in the opposite direction to the main effect, as is the case in Langer et al. (2018), is statistically possible. However, because in this case the indirect effect does not conceptually explain the main effect, the indirect effect may be an artefact that occurred because of a spurious correlation between the mediators (i.e., information known and open treatment) and dependent variable (i.e., organisational attractiveness) (Fiedler et al., 2018; Yzerbyt et al., 2018).

These findings show that applicants perceive organisations which use AI in the hiring process as less attractive than those using human hiring. Due to this, AI is perceived to be worse for organisational attractiveness. Although in Langer, Konig, & Fitili (2018), AI is not compared to human hiring, it shows that the level of transparency one has surrounding the hiring process may influence how attractive it is perceived to be. In this case, the negative organisational perceptions may be alleviated through giving applicants more information on how AI hiring works.

4.5.3 Perceptions of use

Three studies assess the perceptions of the usability of AI hiring technologies (Suen et al., 2019; Kodapanakkal et al., 2020; Oberst et al., 2020). See Table 7 for an overview of the studies.

Table 7 Papers assessing perceptions of use of AI hiring

Citation	Key finding(s)	HIRE evaluation
Suen et al. (2019)	Applicants showed less favourability to automated interview conditions than the human interview condition	Worse
Kodapanakkal et al. (2020)	People indicate they are more likely to adopt AI hiring when outcomes are favourable	Unclear
Oberst et al. (2020)	Recruiters use human candidate recommendations to a greater extent than algorithmic candidate recommendations	Worse

Table 8 Papers assessing emotional perceptions of AI hiring

Citation	Key finding(s)	HIRE evaluation
Lee (2018)	Algorithmic hiring decisions rated as less trustworthy and more negative compared to human hiring decisions	Worse
Langer et al. (2019)	AI interviews result in greater ratings of creepiness than human interviews	Worse

Suen et al. (2019) found that applicants in an interview featuring an AI decision agent have less favourability towards this process than if the agent is a human ($np^2 = 0.391$, $p < 0.001$), and this effect is large. Favourability, meaning how beneficial the outcome is to the individual, was measured with 10 questions adopted from Guchait et al. (2014) on a 5-point Likert scale. Favourability was also found to be indicative of the decision to adopt or reject AI hiring technology in Kodapanakkal et al.'s (2020) experiment. Favourability was manipulated by stating whether it increases or decreases the chance of someone finding employment. When participants were given the choice regarding usage of AI hiring technology, they were more likely to embrace the technology if outcome favourability was high ($z = 12.26$, $p < 0.001$).

Oberst et al. (2020) extend these studies by looking at perceptions of recruitment professionals, and find that they greatly prefer using human judgements than algorithmic assessments in candidate selection decisions. They assessed how recruitment professionals make decisions in a fictitious scenario about candidate selection when they are given information from an algorithm regarding the candidate's suitability for a job with three levels; "sufficient", "satisfactory", and "good". They were also given a co-worker's recommendation on each candidate with three levels; "I recommend this person totally", "this person causes an excellent impression", and "this person does not inspire confidence". Results assessed to what extent the recruiters used the algorithmic assessment and human judgements in their selection decisions, giving each factor an average utility score. Bayesian hierarchical analysis revealed that the average utility assigned to co-worker's recommendation ($M = 58.69$, $SD = 9.47$) was higher than the algorithm ($M = 22.34$, $SD = 11.23$). Thus, there is a large effect ($d = 3.49$) of recruiters using co-worker's recommendations in selection decisions to a greater extent than an algorithmic assessment.

Again, here we see that perceptions around the usability of AI are worse than those of human hiring. This reflects a possible obstacle in the adoption of AI hiring. As shown in Kodapanakkal et al. (2020), favourability of the hiring outcome was indicative of the decision to adopt AI. Thus, these negative perceptions surrounding usability may be driven by a fear of poor hiring outcomes.

4.5.4 Emotional perceptions

Candidate's emotional perceptions which were evoked by AI hiring is assessed in two studies (Lee, 2018; Langer, Konig & Papathanasiou, 2019). See Table 8 for an overview of the studies.

Lee (2018) found that people trust humans much more than algorithms in hiring decisions, and have slightly more negative feelings towards AI hiring. They measured trust of decision process by having participants answer on a 7-point Likert scale and emotional response by asking questions adapted from Larsson, 1987; Weiss et al., 1999 on a 7-point Likert scale. They found a large effect that people trust algorithms less than humans during

hiring ($d = 0.951$, $p < 0.0001$), and a small effect that they have more negative feelings towards algorithmic hiring ($d = 0.394$, $p < 0.05$). In addition, Langer et al. (2019) measured creepiness of algorithmic hiring from Langer and Konig (2018) on a 7-point scale. They found a medium effect that automated interviews evoked feelings of "creepiness" ($np^2 = 0.06$, $p < 0.01$).

These studies show preliminary evidence that people have negative connotations surrounding the AI hiring process, but that the size of the effect may vary based on the type of emotional perception, the largest effect seen in the lack of trust in AI hiring. Thus, emotional perceptions are worse for AI hiring than for human hiring.

4.5.5 Additional perceptions

Five experiments have looked at the role of consistency in perceptions surrounding algorithmic hiring (Langer et al. 2018, 2019; Kaibel et al. 2019; Acikgoz et al. 2020; Noble et al. 2021). See Table 9 for an overview of the studies.

Kaibel et al. (2019) hypothesized that algorithms in the context of hiring would be viewed as more consistent than a human decision-making process. They measured consistency on a 5-point Likert scale developed by Bauer et al. (2001). Although initially they found a large effect ($d = 1.16$) of algorithmic hiring ($M = 4.36$, $SD = 0.67$) being rated as more consistent than humans ($M = 3.36$, $SD = 1.02$), it could not be replicated in the second study which included additional contextual information, due to scores being near the end-points of the scale (algorithms- $M = 4.51$, $SD = 0.60$, humans- $M = 4.44$, $SD = 0.66$). Furthermore, the two other studies (Langer, Konig & Papathanasiou, 2019; Langer, Konig & Fitili, 2018) which measured consistency on a 5-point Likert scale adapted from Bauer et al. (2001) and Warstza (2012) could not find any significant effect of consistency ($np^2 = 0.00$, $p > 0.05$), even when varying stakes ($np^2 = 0.03$, $p > 0.05$) and level of information ($np^2 = 0.00$, $p > 0.05$). However, the last two studies (Acikgoz et al., 2020; Noble, Foster, & Craig, 2021) found that consistency, also measured using Bauer et al.'s (2001) items via 5- or 7-point Likert scales, was rated significantly higher in AI hiring processes than human hiring, with an effect size ranging from small to large (d ranged from 0.36 to 0.90, $p < 0.001$). Thus, there is mixed evidence concerning the relationship between consistency and type of hiring. Considering that the studies used the same measure of consistency, and that effect sizes varied substantially, it is likely that the ratings were sensitive to the experimental methodology factors.

Lastly, a few additional negative perceptions surrounding algorithmic hiring were discovered. Kaibel et al. (2019) measured personableness of the hiring process through a four item scale adapted from Wilhelmy et al. (2019) and found a medium to large effect ($d = 0.618$-0.904, $p < 0.001$) that the AI selection process is less personable. Langer, Konig & Papathanasiou (2019) measured perceived behavioural control on a 7-point Likert scale adapted from Langer et al. (2017). They found that perceived behavioural control was lower on automated interviews ($np^2 = 0.10$, $p < 0.01$), with a moderate effect size. Finally, Newman et al. (2020) measured decontextualisation—the ability of the algorithm to accurately combine and weigh pieces of information—using a 7-point Likert scale and found a medium effect that algorithms are perceived to have more decontextualisation in high transparency conditions ($d = 0.53$, $p = 0.006$).

These last findings show additional perceptual areas where AI is considered worse than humans. Thus, perceptions regarding AI hiring extend beyond the ethical, organisational,

Table 9 Papers assessing additional perceptions of AI hiring

Citation	Key finding(s)	HIRE evaluation
Langer et al. (2018)	AI interviews have no difference in ratings of consistency when varying stakes or level of information	Unclear
Kaibel et al. (2019)	Algorithmic hiring initially found to be perceived as more consistent than human hiring, however, this was not replicated in follow-up study	Unclear
Acikgoz et al. (2020)	AI hiring rated as more consistent than human hiring	Better
Noble et al. (2021)	AI hiring rated as more consistent than human hiring	Better
Kaibel et al. (2019)	AI hiring rated as less personable than human hiring	Worse
Langer et al. (2019)	AI interviews rated as lower on perceived behavioural control than human interviews	Worse
Langer et al. (2019)	AI interviews and human interviews had no difference in ratings of consistency	Equal
Newman et al. (2020)	Algorithms for hiring perceived to have more decontextualisation in high transparency condition than human methods of hiring	Worse

Table 10 Overview of Findings for each Theme Evaluated with the HIRE Framework

Theme	HIRE evaluation			
	Better	Equal	Worse	Unclear
Efficiency	2	2	0	0
Performance	2	0	0	0
Diversity	5	0	1	2
Perceptions	4	3	15	6

Values indicate the number of papers falling under each HIRE evaluation category per theme

usage, and emotional. Further research findings may distinguish even more perceptual differences between AI and human hiring.

5 Discussion

The purpose of this paper was to propose the HIRE framework that can be used to evaluate whether AI is better, equal, or worse than humans in studies that examine how AI can be used in the recruitment process. This framework can be used in the literature evaluating AI and recruitment going forward to allow for easier comparison across studies. Overall the HIRE framework puts emphasis on whether AI is making equally good or better decisions than human recruiters. In the present article, we demonstrate the HIRE framework by conducting a systematic literature on the topic of AI hiring. In this paper we outline experimental results pertaining to the efficiency, performance, diversity, and perceptions of algorithmic hiring in line with the proposed HIRE framework. We report the HIRE framework evaluations for each study below in Table 10. We hope to inspire other authors to use the same classifications.

In the efficiency theme, we found that AI was the same or better than human hiring. AI was able to simulate human decisions (Naim et al. 2018; Stein 2018), predict human hiring outcomes even better than another human (Bergman et al. 2020), or increase the fill rates of certain job positions (Horton, 2017). Due to this, AI can be used in firms wanting to improve efficiency of the hiring process. In the cases where AI can simulate human hiring, these decisions will be more rapid due to the data driven nature of AI (Fernandez-Loria et al. 2020). In fact, three prominent AI hiring platforms feature claims that hiring using their tools reduce hiring time by as much as 70–90% (Ideal 2021; Hirevue 2021; Pymetrics 2021). Thus, AI can be used to make hiring decisions which are similar to those that would be made by a human, and because these AI decisions are rapid, this can produce time cost-savings in the hiring process.

The second theme we assessed was performance, meaning the ability of AI to hire candidates which will have better on the job performance outcomes. Although research in this area is currently limited, the evidence suggests AI is better at hiring candidates with more favourable job performance outcomes (Sajjadiani et al. 2019; Bergman et al. 2020). This preliminary evidence shows that although AI has limited abilities to predict job performance to a large degree, it is still an improvement over human hiring. Thus, for firms wanting to hire candidates who will have better job performance, they should take caution in using AI and evaluate findings for effectiveness.

In the diversity theme, most evidence points towards AI being better than human hiring. This is in line with prior literature which proposes that AI can eradicate human biases which cause discrimination and unequal outcomes for certain groups (Sanchez-Monedero et al. 2020; Raghaven et al. 2020). However, it has also been proposed that AI can be designed to perpetuate human biases, and thus create worse outcomes for minority groups (Vasconcelos, Cardonha, & Goncalves, 2018; Yarger, Payton, & Neupane 2019). We find that both these situations are possible, depending on the type of AI (Bergman, Li, & Raymond, 2020). Static supervised learning models had worse diversity outcomes than humans, whereas upper confidence bound models had better diversity outcomes. We also find evidence of situations where AI is designed to be fair, but discriminates on the basis of external factors (Chen et al., 2018; Lambrecht & Tucker, 2019). However, in these studies the AI is not compared to humans so it is unclear whether the algorithms are an improvement from status-quo hiring methods. Finally, we find that the diversity-validity dilemma can occur when algorithms are designed to be more fair (Allred, 2019; Song, 2018), meaning that increased diversity decisions come at the cost of lower validity of the algorithm, but that algorithmic regularization techniques can overcome this. For organisations wanting to improve diversity in their hiring outcomes through AI, both the type of AI and the external factors should be closely examined for aspects which may bias the decisions in favour of certain groups.

Lastly, we explored literature surrounding how AI is perceived. Aside from the mixed findings for fairness and consistency, AI hiring was perceived as worse than human hiring on every other perception outcome including privacy concern (Langer et al. 2019), organisational attractiveness (Kaibel et al., 2019; Langer, Konig & Fitili 2018; Langer et al. 2019), favourability (Suen et al., 2019), trust (Lee, 2018; Oberst et al., 2020), emotional response (Lee, 2018), creepiness (Langer, et al. 2019), personableness (Kaibel et al., 2019), perceived behavioural control (Langer et al. 2019), and decontextualisation (Newman et al., 2020). These effects were found to be subject to additional factors such as the individual difference of perceived uniqueness for organisational attractiveness (Kaibel et al., 2019), and level of transparency for decontextualisation (Newman et al., 2020). Thus, organisations which adopt AI in their hiring process should be conscious to the reactions of applicants.

As a whole, these findings reflect a dichotomy that despite evidence that AI is mostly better for efficiency, performance, and diversity, perceptions towards AI hiring are predominantly worse. This is in line with algorithm aversion (Dietvorst et al., 2015) which is a phenomenon that people prefer human decisions over algorithmic ones, even when the algorithms are shown to be more accurate. People have less trust in algorithms, particularly for subjective decisions (Castelo et al. 2019). Since employee selection can be a subjective process (Highhouse, 2008), this may explain why we found perceptions to be so negative towards AI hiring. The reliance on human judgement over data driven tools reflects a challenge for the adoption of AI hiring. There have been factors identified which can influence the cognitive trust in algorithms, and these include tangibility, transparency, reliability, and immediacy behaviours (Glikson & Woolley, 2020). It is also important to note that in the perception studies, none of the studies described a situation where the hiring decision is made on the basis of a mix of algorithm and human judgement. Since none of the leading AI companies suggest hiring decisions should be purely made by algorithms, and human involvement with algorithms has been found to lower aversion (Jussupow et al. 2020), it is possible that aversion may not be a problem in real world AI hiring scenarios.

5.1 Future research areas

Due to AI hiring research still being in infancy, there are many future research avenues to explore. In line with the HIRE framework, we propose that future research on AI hiring should use human hiring as a comparator. As this was the case for the majority of studies outlined, some findings remain unclear due to the lack of comparator. While the four research themes we outline could all benefit from replication studies, we offer three novel areas to conduct research on this topic.

First, by comparing the AI hiring methods used in the literature reviewed to what is being practically used, we observe an academic-practitioner gap. This means that many of the applied uses for AI hiring is ahead of what is being studied. Specifically, a common method of increasing diversity and fairness in AI hiring has been to remove sensitive variables and those which have group differences, a method known as "blinding" (Hirevue, 2021; Mevitae, 2021). Thus, empirical assessment of such AI hiring tools can enhance conclusions about which types of AI can make better hiring decisions.

Another area of interest would be to look at how hiring outcomes vary depending on the stage of the recruitment process that the algorithms are utilised in. All of the studies outlined in this review focus on one specific area of AI hiring, from how the job advert is shown (Lambrecht & Tucker, 2019) to how candidates are ranked (Suhr, Hilgard, & Lakkaraju, 2020; Chen et al. 2018) to CV/ candidate background analysis (Sajjadiani et al. 2019; Stein, 2018; Bergman, Li, & Raymond, 2020; Allred 2019), and interview analysis (Naim et al. 2018). In order to understand when AI hiring is most beneficial, a comparison of algorithmic outcomes at each stage can inform us.

Finally, we recommend that diversity in AI hiring be studied in the future through an intersectional lens. In the current literature, diversity in hiring is studied with respect to gender and ethnicity in a unidimensional way. Intersectionality takes into account the interactions between such demographic groups (Lutz 2015), and can tell us more about which groups might be discriminated against in AI hiring. For example, even if there are no significant group level differences across gender and ethnicity, there may be a significant group difference for a specific combination of gender/ethnicity. Due to this, AI hiring which takes into account intersectionality may produce better diversity outcomes in hiring.

We also note that this framework can generalise to other contexts where novel AI decision-making techniques are used in lieu of human decisions. Comparing the outcomes of AI and human decision-making in line with the themes of efficiency, performance, diversity, and perceptions will add value in a variety of contexts. In these cases, the authors can replace the "R" of HIRE with the context they are studying.

6 Conclusions

In this study we proposed the HIRE framework, which can be used in studies that aim to evaluate the impact AI has on the recruitment process. The HIRE framework's primary aim is to increase the ease of comparability for studies of this kind. In particular, with respect to gauging whether AI is an adequate, or even better, substitute for humans. We illustrate the simplicity of applying this framework by conducting a systematic review. Our review highlights that AI is equal or better when it comes to efficiency and performance as compared to humans. We also find that AI is mostly better than humans when it comes to

improving diversity. Finally, we demonstrate that there is a perception that AI is inferior to humans. We acknowledge that the studies included in this review are limited in terms of context, and hope that this critique will be abated in the future as more literature comes on stream. We also hope that authors will adopt the HIRE framework when conducting research in this area to allow for easier comparability, by placing the HIRE framework outcome in the abstract.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10462-022-10193-6.

Acknowledgements This work was funded by The Inclusion Initiative at the London School of Economics and Political Science.

Funding This work was funded by The Inclusion Initiative at the London School of Economics and Political Science.

Data Availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflicts of interest or competing interests.

References

Acikgoz Y, Davison KH, Compagnone M, Laske M (2020) Justice perceptions of artificial intelligence in selection. Int J Sel Assess 28(4):399–416

Agarwal, J., Malhotra, N. K., & Kim, S. S. (2004). Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model

Allred, C. M. (2019). Applying a Metaheuristic Algorithm to a Multi-Objective Optimization Problem Within Personnel Psychology

American Psychological Association. (2022). *APA PsycInfo*. American Psychological Association. Retrieved from https://www.apa.org/pubs/databases/psycinfo/

Bauer TN, Truxillo DM, Sanchez RJ, Craig JM, Ferrara P, Campion MA (2001) Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). Pers Psychol 54(2):387–419

Bergman P, Li D, Raymond L (2020) Hiring as exploration. SSRN Journal. https://doi.org/10.2139/ssrn.3630630

Bigman, Y., Gray, K., Waytz, A., Arnestad, M., & Wilson, D. (2020). Algorithmic discrimination causes less moral outrage than human discrimination.

Brockner J, Konovsky M, Cooper-Schneider R, Folger R, Martin C, Bies RJ (1994) Interactive effects of procedural justice and outcome negativity on victims and survivors of job loss. Acad Manag J 37(2):397–409

Castelo N, Bos MW, Lehmann D (2019) Let the Machine Decide: When Consumers Trust or Distrust Algorithms. NIM Marketing Intelligence Review 11(2):24–29

Chen, L., Ma, R., Hannák, A., & Wilson, C. (2018, April). Investigating the impact of gender on rank in resume search engines. In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1–14).
Cohen J (1962) The statistical power of abnormal-social psychological research: a review. Psychol Sci Public Interest 65(3):145
Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. Hillsdle
Colquitt JA (2001) On the dimensionality of organizational justice: A construct validation of a measure. J Appl Psychol 86(3):386–400. https://doi.org/10.1037/0021-9010.86.3.386
Conlon DE, Porter CO, Parks JM (2004) The fairness of decision rules. J Manag 30(3):329–349
Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144(1):114
Fernández-Loría, C., Provost, F., & Han, X. (2020). Explaining Data-Driven Decisions made by AI Systems: The Counterfactual Approach. arXiv e-prints, arXiv-2001.
Korn Ferry. (2018, January 18). Korn Ferry Global Survey: Artificial Intelligence (AI) reshaping the role of the recruiter. Retrieved May 12, 2021, from https://www.kornferry.com/about-us/press/korn-ferry-global-survey-artificial-intelligence-reshaping-the-role-of-the-recruiter
Fiedler K, Harris C, Schott M (2018) Unwarranted inferences from statistical mediation tests–An analysis of articles published in 2015. J Exp Soc Psychol 75:95–102
Ganguly AR, Gupta A, Khan S (2005) Data mining technologies and decision support systems for business and scientific applications. Encyclopedia of Data Warehousing and Mining. https://doi.org/10.4018/978-1-59140-557-3.ch045
Gartner. (2019). Gartner survey Shows 37 percent of organizations have Implemented AI in some form. Retrieved May 12, 2021, from https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have
Geetha R, Bhanu SRD (2018) Recruitment through artificial intelligence: a conceptual study. Int J Mech Eng Technol 9(7):63–70
Glikson E, Woolley AW (2020) Human trust in artificial intelligence: Review of empirical research. Acad Manag Ann 14(2):627–660
Guchait P, Ruetzler T, Taylor J, Toldi N (2014) Video interviewing: A potential selection tool for hospitality managers–A study to understand applicant perspective. Int J Hosp Manag 36:90–100
Haenlein M, Kaplan A (2019) A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. Calif Manage Rev 61(4):5–14
Highhouse S (2008) Stubborn reliance on intuition and subjectivity in employee selection. Ind Organ Psychol 1(3):333–342
Highhouse S, Lievens F, Sinar EF (2003) Measuring attraction to organizations. Educ Psychol Measur 63(6):986–1001
HireVue. (2021). HireVue: Video interview software & recruitment platform. Retrieved May 12, 2021, from https://www.hirevue.com/
Holmes, A. (2019, October 8). *Ai could be the key to ending discrimination in hiring, but experts warn it can be just as biased as humans*. Business Insider. Retrieved January 24, 2022, from https://www.businessinsider.com/ai-hiring-tools-biased-as-humans-experts-warn-2019-10?r=US&IR=T
Horton JJ (2017) The effects of algorithmic labor market recommendations: Evidence from a field experiment. J Law Econ 35(2):345–385
Ideal. (2021, April 22). Why Ideal?: Talent intelligence system. Retrieved May 12, 2021, from https://ideal.com/why-ideal/
Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion. In ECIS.
Kaibel, C., Koch-Bayram, I., Biemann, T., & Mühlenbock, M. (2019, July). Applicant perceptions of hiring algorithms-uniqueness and discrimination experiences as moderators. In Academy of Management Proceedings (Vol. 2019, No. 1, p. 18172). Briarcliff Manor, NY 10510: Academy of Management.
Kallem SR (2012) Artificial intelligence algorithms. IOSR J Computer Engineering (IOSRJCE) 6(3):1–8
Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research, 1–54.
Kodapanakkal RI, Brandt MJ, Kogler C, Van Beest I (2020) Self-interest and data protection drive the adoption and moral acceptability of big data technologies: A conjoint analysis approach. Comput Hum Behav 108:106303
Konovsky MA, Folger R (1991) The effects of procedures, social accounts, and benefits level on victims' layoff reactions. J Appl Soc Psychol 21(8):630–650
Kuncel NR, Klieger DM, Ones DS (2014) In hiring, algorithms beat instinct. Harv Bus Rev 92(5):32–32

Lambrecht A, Tucker C (2019) Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. Manage Sci 65(7):2966–2981
Langer M, König CJ (2018) Introducing and testing the Creepiness of Situation Scale (CRoSS). Front Psychol 9:2220. https://doi.org/10.3389/fpsyg.2018.02220
Langer M, König CJ, Fitili A (2018) Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. Comput Hum Behav 81:19–30
Langer M, König CJ, Papathanasiou M (2019) Highly automated job interviews: Acceptance under the influence of stakes. Int J Sel Assess 27(3):217–234
Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings, Volume 25, Issue 4.
Larsson G (1987) Quick measurement of individual stress reaction level: Development of the Emotional Stress Reaction Questionnaire (ESRQ). Report from Defense Technical Information Center.
Laugwitz, B., Held, T., & Schrepp, M. (2008, November). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group (pp. 63–76). Springer Berlin Heidelberg.
Lee MK (2018) Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data Soc 5(1):2053951718756684
Lohr, S. (2021, December 8). *Group backed by top companies moves to combat A.I. Bias in hiring*. The New York Times. Retrieved January 24, 2022, from https://www.nytimes.com/2021/12/08/technology/data-trust-alliance-ai-hiring-bias.html
Lutz, H. (2015). Intersectionality as method. DiGeSt. Journal of Diversity and Gender Studies, 2(1–2), 39–44.
MeVitae. (2021). Bias free hiring within your ats. Retrieved May 12, 2021, from https://www.mevitae.com/
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097
Naim, I., Tanveer, M. I., Gildea, D., & Hoque, M. E. (2018). Automated Analysis and Prediction of Job Interview Performance. IEEE Transactions on Affective Computing, 9(02):191204.
Newman DT, Fast NJ, Harmon DJ (2020) When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. Organ Behav Hum Decis Process 160:149–167
Nica E, Miklencicova R, Kicova E (2019) Artificial intelligence-supported workplace decisions: Big data algorithmic analytics, sensory and tracking technologies, and metabolism monitors. Psychosociol Issues Hum Resour Manag 7(2):31–36
Noble SM, Foster LL, Craig SB (2021) The procedural and interpersonal justice of automated application and resume screening. Int J Select Assess 29:139–153
Oberst, U., De Quintana, M., Del Cerro, S., & Chamarro, A. (2020). Recruiters prefer expert recommendations over digital hiring algorithm: a choice-based conjoint study in a pre-employment screening scenario. Management Research Review.
Parikh, N. (2021, December 10). *Understanding bias in AI-enabled hiring*. Forbes. Retrieved January 24, 2022, from https://www.forbes.com/sites/forbeshumanresourcescouncil/2021/10/14/understanding-bias-in-ai-enabled-hiring/?sh=53df8c7c7b96
Pymetrics. (2021). Solutions. Retrieved May 12, 2021, from https://www.pymetrics.ai/solutions
Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 469–481).
Roth PL, Bevier CA, Bobko P, SWITZER III, F. S., & Tyler, P. (2001) Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. Pers Psychol 54(2):297–330
Sajjadiani S, Sojourner AJ, Kammeyer-Mueller JD, Mykerezi E (2019) Using machine learning to translate applicant work history into predictors of performance and turnover. J Appl Psychol 104(10):1207
Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020, January). What does it mean to'solve'the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 458–468).
Smith HJ, Milberg SJ, Burke SJ (1996) Information privacy: Measuring individuals' concerns about organizational practices. MIS Q 20:167–196
Song, Q. (2018). Diversity shrinkage of Pareto-optimal solutions in hiring practice: Simulation, shrinkage formula, and a regularization technique (Doctoral dissertation, University of Illinois at Urbana-Champaign).
Spar, B., Pletenyuk, I., Reilly, K., & Ignatova, M. (2018). Global Recruiting Trends 2018(Rep.). Retrieved May 12, 2021, from LinkedIn website: https://news.linkedin.com/2018/1/global-recruiting-trends-2018

Stein, S. K., Goldberg, A., & Srivastava, S. B. (2018). Distinguishing Round from Square Pegs: Predicting Hiring Based on Pre-hire Language Use (No. repec: ecl: stabus: 3627).
Suen HY, Chen MYC, Lu SH (2019) Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? Comput Hum Behav 98:93–101
Sühr, T., Hilgard, S., & Lakkaraju, H. (2020). Does Fair Ranking Improve Minority Outcomes? Understanding the Interplay of Human and Algorithmic Biases in Online Hiring. arXiv preprint arXiv:2012.00423.
Sunstein CR, Kahneman D, Schkade D (1998) Assessing punitive damages. Yale Law Journal 107(50):2071–2153
Upadhyay AK, Khandelwal K (2018) Applying artificial intelligence: implications for recruitment. Strateg HR Rev 17:255–258
Vasconcelos, M., Cardonha, C., & Gonçalves, B. (2018, December). Modeling epistemological principles for bias mitigation in AI systems: an illustration in hiring decisions. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 323–329).
Warrenbrand, M. (2021). Applicant justice perceptions of machine learning algorithms in personnel selection.
Warszta, T. (2012). Application of Gilliland's model of applicants' reactions to the field of web-based selection(Doctoral dissertation, Christian-Albrechts Universität Kiel).
Weiss HM, Suckow K, Cropanzano R (1999) Effects of justice conditions on discrete emotions. J Appl Psychol 84(5):786
Wilhelmy A, Kleinmann M, Melchers KG, Lievens F (2019) What do consistency and personableness in the interview signal to applicants? Investigating indirect effects on organizational attractiveness through symbolic organizational attributes. J Bus Psychol 34(5):671–684
Yarger L, Payton FC, Neupane B (2019) Algorithmic equity in the hiring of underrepresented IT job candidates. Online Inf Rev 44:383–395
Yzerbyt V, Muller D, Batailler C, Judd CM (2018) New recommendations for testing indirect effects in mediational models: The need to report and test component paths. J Pers Soc Psychol 115(6):929