UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

AI and discriminative decisions in recruitment : Challenging the core assumptions

Seppälä, Päivi; Maecka, Magdalena

2024-03-30
SAGE Publications Ltd
http://hdl.handle.net/10138/574792
Seppälä, P & Maecka, M 2024, 'AI and discriminative decisions in recruitment : Challenging the core assumptions', Big Data & Society, vol. 11, no. 1, pp. 1-12. https://doi.org/10.1177/20539517241235872

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Check for updates

Original Research Article

BIG DATA & SOCIETY

Big Data & Society
January–March: 1–12
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517241235872
journals.sagepub.com/home/bds
Sage

AI and discriminative decisions in recruitment: Challenging the core assumptions

Päivi Seppälä[1] **and Magdalena Małecka**[2]

Abstract

In this article, we engage critically with the idea of promoting artificial intelligence (AI) technologies in recruitment as tools to eliminate discrimination in decision-making. We show that the arguments for using AI technologies to eliminate discrimination in personnel selection depend on presuming specific meanings of the concepts of rationality, bias, fairness, objectivity and AI, which the AI industry and other proponents of AI-based recruitment accept as self-evident. Our critical analysis of the arguments for relying on AI to decrease discrimination in recruitment is informed by insights gleaned from philosophy and methodology of science, legal and political philosophy, and critical discussions on AI, discrimination and recruitment. We scrutinize the role of the research on cognitive biases and implicit bias in justifying these arguments – a topic overlooked thus far in the debates about practical applications of AI. Furthermore, we argue that the recent use of AI in personnel selection can be understood as the latest trend in the long history of psychometric-based recruitment. This historical continuum has not been fully recognized in current debates either, as they focus mainly on the seemingly novel and disruptive character of AI technologies.

Keywords

AI, discrimination, recruitment, cognitive bias, implicit bias, fairness, objectivity, rationality

Introduction

'Unlike humans, technology is blind to gender, ethnicity, age and background. Our bias-free skill assessments can automatically be scored while analyzing an endless amount of data to help the hiring team locate the best candidates in no time', a company called Canditech assures its clients (Canditech, 2022). The company designs artificial intelligence (AI) technology which assists recruiters in selecting employees for job openings. Another company, HireVue, guarantees that it will find the best-performing employees by evaluating 'each candidate in a large pool quickly and fairly, so you can make high-quality inclusive hiring decisions' (HireVue, 2022). For some, the promises of firms such as Canditech or Hirevue may resemble pages from a dystopian sci-fi novel. In fact, such a way of thinking about recruitment and AI is a reality in organizations worldwide (Raghavan et al., 2020). Consequently, the ethical aspects of these new recruitment practices have become an important topic in academic discussions about AI (Dennis and Aizenberg, 2022).

In this article, we engage critically with the idea of promoting AI technologies[1] in recruitment as tools to eliminate discrimination[2] in decision-making. To this effect, we took a closer look at justifications for introducing AI technologies into organizations which attempt to reduce discrimination during the phase of recruitment called 'personnel selection'.[3] We analyzed a body of literature ranging from research in the fields of AI marketing, human resources management (HRM), computer science, AI ethics to academic and popular research on cognitive psychology. Based on this analysis, we identified the justifications presented by the proponents of AI solutions to discrimination in recruitment. In short, they claim that human decision-making is flawed and often leads to discrimination. Therefore, AI solutions are needed to secure objectivity and fairness in decision-making. Furthermore, they claim that the potential biases of the current machine learning systems, which may also lead to unfair and discriminatory results, can be corrected by technical means within AI

[1]Practical Philosophy, University of Helsinki, Helsinki, Finland
[2]Aarhus Institute for Advanced Studies, Aarhus University, Aarhus, Denmark

Corresponding author:
Magdalena Małecka, Aarhus University, Aarhus, Denmark.
Email: malecka.magdalena@gmail.com

Creative Commons NonCommercial-NoDerivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NoDerivs 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits any use, reproduction and distribution of the work as published without adaptation or alteration, provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

development. In scrutinizing these claims, we show that the arguments for using AI technologies, such as machine learning, to eliminate discrimination in personnel selection depend on presuming specific meanings of the concepts of rationality, bias, fairness, objectivity and AI.

We contribute to the existing literature on AI and discrimination in recruitment in four ways. First, we show that a specific understanding of AI and its role in recruitment is shared across diverse scholarly and practical contexts – in the AI industry and its development, in policy reports and proposals, in research on cognitive and implicit biases and in recruitment practices. Second, we provide an original synthesis of some aspects of critical literature on AI in recruitment. We draw together dispersed bodies of this literature and show that bringing the existing critical points into dialogue is fruitful. In the case of our analysis, it allows us to uncover and scrutinize the presumptions about basic concepts featured in justifications for using AI in personnel selection. Third, and related to the previous point, we emphasize that the literature in philosophy of science which discusses the limitations of behavioral research on biases offers an important critical angle for unpacking and questioning the justifications for using AI in recruitment. This literature has been overlooked thus far in the debates about practical applications of AI. We believe that it should be given more prominence in critical reflections on this topic. Fourth, we argue that the recent use of AI in personnel selection can be understood as the latest trend in the practices of psychometric-based recruitment. Therefore, many challenges and perils of AI-based recruitment recently discussed in critical literature reviews are similar to the problems with psychometric recruitment tools. This has not been fully recognized in current debates, which focus mainly on the seemingly novel and disruptive character of AI technologies.

Justifications for relying on AI in personnel selection: an overview

We start by analyzing the literature which discusses attempts to reduce discrimination in personnel selection by relying on AI technology. This allows us to identify several claims, shared across the fields of HRM, computer science and AI, as well as in the behavioral sciences, that justify the use of AI in personnel selection.

An umbrella review of the literature in HRM, computer science and AI

We conducted an umbrella review (Grant and Booth, 2009) of 17 review articles and a review report from journals in the fields of HRM, computer science and AI. (We explain in detail the method we used for selecting the articles in Appendix 1.) Our umbrella review covers the existing reviews of the literature on AI-based discrimination in personnel selection in these fields and it aims to fill the gap in the review literature by providing the reader with an overview of the justifications for AI-based personnel selection. The limitation of focusing on review articles rather than original research articles, is that umbrella reviews might lose some detail, such as specific arguments and topics that one might find when analyzing individual research articles (Grant and Booth, 2009). However, because our aim was to analyze claims made in various types of literature to justify the use of AI in personnel selection, we consider an umbrella review as a good methodological tool to identify claims shared across research fields.

In general, the reviews we analyzed acknowledge that AI has the potential to address discrimination in personnel selection. AI-based personnel selection is considered to be fairer and more objective compared to decisions about hiring based on human judgment. AI systems are seen as ensuring that job applicants are evaluated according to their skills or performance, instead of on their membership in a socially salient group (Bailao Goncalves et al., 2022; Chilunjika et al., 2022; Köchling and Wehner, 2020; Will et al., 2023). The idea of AI-based personnel selection overcoming human biases is present in the AI industry's marketing (Bogen and Rieke, 2018; Drage and Mackereth, 2022; Raghavan et al., 2020; Sánchez-Monedero et al., 2020), HRM literature (Bailao Goncalves et al., 2022; Chilunjika et al., 2022; Cho et al., 2023; Köchling and Wehner, 2020; Pessach and Shmueli 2022; Will et al., 2023), as well as in the discussion on AI ethics (Hunkenschroer and Luetge, 2022).

The concept of human bias appears in two kinds of claims. First, there are *essentialist claims* about the nature of human cognition or human decision-making as being biased or irrational (Chilunjika et al., 2022; Cho et al., 2023; Drage and Mackereth, 2022; Fernandes França et al., 2023; Hunkenschroer and Luetge, 2022; Köchling and Wehner, 2020; Pessach and Shmueli, 2022; Raghavan et al., 2020; Tursunbayeva et al., 2022; Will et al., 2023). For instance, Chilunjika et al., (2022: 7) claim that 'the human interface' is 'normally susceptible to biased judgement'. Second, there are *causal claims* about human biases leading to discrimination in personnel selections such as '[s]ome employers have tended to discriminate against women and ethnic minorities, albeit unconsciously' (Hofeditz et al., 2022: 145; see also: Birzhandi and Cho, 2023; Bogen and Rieke, 2018; Chilunjika et al., 2022; Cho et al. 2023; Drage and Mackereth, 2022; Fernandes França et al., 2023; Hunkenschroer and Luetge, 2022; Kaur and Kaur, 2022; Köchling and Wehner, 2020; Nadeem et al., 2021; Pessach and Shmueli, 2022; Raghavan et al., 2020; Sánchez-Monedero et al., 2020; Tursunbayeva et al., 2022; Will et al., 2023).

However, it is also claimed that not only humans are prone to bias and discrimination: AI-based personnel selection also suffers from *algorithmic bias*,[4] namely the fact that biased datasets or biased algorithms make AI systems discriminate against socially salient groups or members of these groups (Birzhandi and Cho, 2023; Bogen and Rieke, 2018; Cho et al. 2023; Fernandes França et al., 2023; Hofeditz et al., 2022; Hunkenschroer and Luetge, 2022; Köchling and Wehner, 2020; Nadeem et al., 2021; Pessach and Shmueli, 2022; Raghavan et al., 2020; Tursunbayeva et al., 2022). Therefore, algorithmic bias is seen as undermining the objectivity and/or fairness of the predictions that AI-based personnel selection systems produce about job applicants (Birzhandi and Cho, 2023; Cho et al. 2023; Fernandes França et al., 2023; Hofeditz et al., 2022; Hunkenschroer and Luetge, 2022; Pessach and Shmueli, 2022; Köchling and Wehner, 2020). Nevertheless, most of the articles are positive about the possibility to address these challenges by AI developers incorporating new technical solutions (Bailao Goncalves et al., 2022; Birzhandi and Cho, 2023; Hofeditz et al., 2022; Hunkenschroer and Luetge, 2022; Kaur and Kaur, 2022; Pessach and Shmueli, 2022; Raghavan et al., 2020; Will et al., 2023).

Analysis of the literature on AI and behavioral science

The reviews we analyzed make explicit references to the notions of cognitive biases (Cho et al., 2023; Köchling and Wehner, 2020; Will et al., 2023) and implicit biases (Drage and Mackereth, 2022). They rely on the results of the research in cognitive and social psychology to indicate the sources of the 'human bias' that may lead to discrimination.[5] Therefore, we scrutinized what the most prominent scholars of cognitive bias and implicit bias have written on AI as a solution to discrimination in personnel selection. We also note that these scholars have promoted such AI solutions in some of their recent publications. For instance, Daniel Kahneman, one of the leading researchers in behavioral economics and cognitive psychology of judgment and decision-making has supported the introduction of AI to correct the errors of human decision-making (Kahneman, 2011). Another prominent scholar of decision-making and the co-author of the bestseller, *Nudge* (Thaler and Sunstein, 2008), Cass Sunstein, has hailed the use of AI to eliminate human biases (Kleinberg et al., 2018; Sunstein, 2019). Finally, Anthony Greenwald, one of the main theoreticians of implicit bias research (Greenwald and Banaji, 1995) as well as the co-developer of the implicit association test (IAT) (Greenwald et al., 1998), has recently mentioned AI as one means of reducing discrimination caused by decision-makers' implicit biases (Greenwald et al., 2022).

The ideas of these scholars are similar to the ones we identified in the review articles. For instance, in their recently published popular book, *Noise*, Kahneman joined forces with Cass Sunstein and Oliver Sibony to give advice to decision-makers on how to overcome both bias and noise, the main sources of error in human judgment. Kahneman et al. (2021) argue that bias can be eliminated or reduced by algorithms, even though algorithms themselves may be biased due to the features of their design and the type of data that feeds them. Nevertheless, the authors claim that 'an algorithm can be more accurate (…) while producing less racial discrimination than human beings do' (2021: 335), particularly, in the context of personnel selection. They propose 'an inescapable conclusion: although a predictive algorithm in an uncertain world is unlikely to be perfect, it can be less imperfect than noisy and often-biased human judgment' (2021: 336).

The justificatory claims for AI-based personnel selection

Based on the umbrella review of the above-mentioned literature on HRM, computer science and AI, as well as the analysis on the recent writings of the prominent researchers of cognitive bias and implicit bias we identified the following set of claims which justify the use of AI in personnel selection:

1. Human cognition and decision-making are irrational and biased.
2. Biases lead to flawed judgment and decision-making which in the context of personnel selection means unfair and discriminative decision-making.
3. Personnel selection should be fair and objective: based on assessing applicants' skills, merit and performance which are the only relevant selection criteria.
4. AI-based personnel selection is a technology that guarantees fairer and more objective decision-making, and it provides a contrast to the flaws and biases of human decision-making.
5. Even though current AI-based solutions might also contain technical biases, they can be tackled by technological means through further AI development.

We conclude that the justificatory claims are widely shared across diverse research and practice communities, although they are not always discussed together in the way we reconstruct them here as a target of our critical discussion.[6]

Criticism

In the below sections, we scrutinize the identified claims, and we argue that they presume specific meanings of the concepts, such as rationality, bias, fairness, objectivity and AI. We show that claim 1 is motivated by presuming a particular meaning of the concept of rationality and that claim 2 is based on a questionable understanding of the causal efficacy of human biases. Claims 3 and 4 depend

on specific meanings of the concepts of fairness and objectivity, while claim 5 is based on the view of AI as a technology detached from its social and political surroundings.

Our criticism is based on insights from philosophy and methodology of science, as well as on points raised by other scholars in critical discussions on AI, discrimination and recruitment, mainly from the fields of legal studies, critical management studies, computer science, sociology and science and technology studies. We stress that apart from providing important insights on the theory and practice of AI-based solutions, bringing these points together allows us to *make explicit* the presumed understanding of concepts, such as bias, rationality or fairness. We show that justifications for relying on AI to address discrimination in recruitment depend on these presumptions, which are, however, contested and highly problematic.

Human cognition and decision making are irrational and biased

The argument that human decision-making is irrational and biased plays an important role in justifying AI-based solutions to discrimination, as expressed in justificatory claims 1 and 2. The proponents of AI solutions often argue that due to irrationality and bias in human judgment and decision-making, personnel selection practices may lead to discriminative outcomes, and often do. AI technology is viewed as being more reliable than biased and irrational humans in processing information about candidates. AI is meant to select those candidates who are predicted to perform best in a job. The idea that human decision-making is biased and that large parts of it operate at a subconscious level is treated as the reliable and established result of the scientific research on cognition. The origin of the idea is often attributed to Amos Tversky and Daniel Kahneman's research on judgment and decision-making in cognitive psychology initiated in the 1970s. This research became very influential across diverse scientific fields. For instance, it inspired the development of behavioral economics (e.g. Thaler, 2000). Another research tradition that has advanced the idea of biased human decision-making is implicit bias research, which originated from the adoption of theories and methodologies of cognitive psychology, such as methodologies of attention research and implicit memory research, to ask questions about the automatic and unconscious workings of human social cognition (e.g. Payne and Gawronski, 2010). Like the theories of Tversky and Kahneman, the research findings of implicit bias captured the popular imagination and entered professional contexts such as policy circles, business and management (Machery, 2022). The popularized versions of cognitive psychology and social psychology also inform discourses and debates about using AI in recruitment practices to overcome biases in human decision-making (Burrell and Fourcade, 2021). However, they are simplifications that hide presumptions of the research and mask important challenges faced by studies on prospect theory, heuristics and biases, dual process theories and implicit bias. We intend to flesh out the ongoing scholarly debates about this research. We argue that they should be treated seriously, as the existing methodological criticism significantly challenges the popular narratives about research in cognitive psychology and therefore hampers the idea of AI as a tool to overcome biases in human decision-making.

We start with Tversky and Kahneman and the question of whether their work supports the view that human decision-making is biased and irrational. Tversky and Kahneman are well known for continuing Herbert Simon's work on bounded rationality, which challenges the view of rational individual decision-making, exemplified in the so-called rational choice theories (particularly, expected utility theory) (e.g. Simon, 1984). They developed prospect theory as an alternative theory of decision-making, claiming that people do not follow decision-making procedures as expected utility theory defines them: people's decisions in most contexts are irrational or boundedly rational. Kahneman, Tversky and their collaborators introduced the notion of heuristics and biases to study judgments under conditions of uncertainty (Kahneman et al., 1982). They argued that heuristics as mental shortcuts or rules of thumb are used to judge the likelihood of events and that reliance on heuristics leads to systematic errors: biases (Tversky and Kahneman, 1974). Kahneman is also a well-known advocate of dual-process theories. He claims that reliance on heuristics engages fast, automatic, intuitive and unconscious processes of the mind, so-called System 1 processes, whereas making decisions or judgments based on reasoning, statistical estimations or hypothetical thinking engages slow, controlled, reflective, rule-based, effortful and conscious cognitive processes, the so-called System 2 (Kahneman, 2011).

It is too often overlooked in the popular narratives about heuristics and biases that this research approach is an offshoot of the so-called cognitive revolution, studying the human mind as if it were a computer (e.g. Edwards, 1996; Miller, 2003). The origins of the prospect theory in the cognitive revolution manifest themselves in the way the theory envisions decision-making – namely as a procedure consisting of steps of information processing (Małecka, 2021). This means that the theory of decision-making proposed by Tversky and Kahneman presumes a view of the mind as an information-processing computer. Hence, it is unclear in which sense theories of decision-making from cognitive psychology, such as prospect theory, model the cognition of 'humans', as stated in the narratives of cognitive psychologists and behavioral economists, who claim that their studies of the behavior of 'biased' people are

'more realistic' and therefore more suitable for informing policy and other practical contexts (Thaler and Sunstein, 2008).

Furthermore, the criticism of rational decision-making advanced by Tversky and Kahneman is ambiguous. On the one hand, they suggested that expected utility theory is a theory which fails to explain people's decisions and behavior; on the other hand, they argued that it can be accepted as a normative theory which sets a standard for rational behavior (Kahneman and Tversky, 1979). Psychologists have long criticized Tversky and Kahneman's research for treating the notion of rationality derived from the expected utility theory as a normative ideal for decision-making (Gigerenzer, 1996). This criticism makes it clear that the claim, which is made in many contexts, that psychological and behavioral research provides evidence about human decision-making being biased or irrational, can be understood only because of interpreting evidence from the point of view of a standard of rationality. It is not self-evident to commit to the standard derived from the expected utility theory and treat it as a yardstick for all human decision-making. In addition, this normative rationality standard is related to the ideal of frictionless computer performance, as theories of rational choice also have affinities with the early research on AI and the algorithmic view of decision-making is behind them (Erickson et al., 2013). In short, the claim that human rationality is bounded and biased depends on accepting such a notion of rationality, as well as the view of cognition as information processing presumed in the research of Tversky and Kahneman and their followers.

Cognitive and implicit biases lead to flawed human judgment and decision-making

Above, we argued that behavior or decision-making can be treated as biased only in light of the normative standard of rationality. However, research on decision-making from the behavioral sciences is claimed not only to demonstrate *that* human decisions and judgments are biased, but also to explain *how* the biases come about. In the context of debates about discrimination in recruitment, the widely shared consensus is that behavioral research, particularly approaches in cognitive and social psychology mentioned in the previous section, provides reliable and robust evidence about cognitive processes which cause flawed judgment, decision-making or other behavior (for instance, and importantly, discrimination) (Storm et al., 2023).[7] The claim is also that these cognitive processes operate mostly at the subconscious level, particularly via the operations of System 1, and result in cognitive or implicit biases.[8]

The concept of bias in psychological and behavioral research, as well as in the narrative based on this research, is rather vague. It is sometimes understood as a cognitive factor which leads to flawed decisions or behaviors (Greenwald and Banaji, 1995), and at other times it is defined as a behavior which is an outcome of cognitive information processing (Tversky and Kahneman, 1974). The term 'cognitive bias' is often used in practical or policy contexts which rely on the results of behavioral research, and it carries this ambiguity – of a cognitive cause and an effect of a cognitive mechanism at the same time. The lack of precision on whether the bias is a cognitive cause or a behavioral effect is not only the result of vagueness on the part of users of scientific research. It originates from the research itself. In social psychology, particularly, the term 'implicit bias' has been used to refer both to an assumed cognitive-level construct and behavioral responses observed in implicit bias measurements, and there is disagreement as to which understanding of the concept should be applied by researchers (Houwer et al., 2013).

One of the reasons why the relationship between cognitive processes and behavioral effects is unclear may be that the theories in cognitive psychology theories like prospect theory and dual system theory, are not mechanistic theories (Eronen, 2020; Grayot, 2020). They are abstract functional theories, which make claims about functions of cognitive processes but do not identify the causal mechanisms which bring about cognitive processes or behaviors. As already noted, these theories have their origins in the so-called cognitive revolution, which defines cognitive processes through an analogy with a computer, computing functions and information processing. There are several decades of discussion on whether information processing procedures or operations proposed by theories such as prospect theory can be interpreted causally, as well as whether the information processing, algorithmic or computational structures studied in cognitive psychology relate to processes in brain structures (e.g. Marr and Poggio, 1976). Moreover, the dual process theories seem to be underdetermined by evidence and there is no evidential support for identifying System 1 and System 2 with a neural architecture (Keren 2013; Grayot, 2020).

The widespread practical applications of research on heuristics and biases, and on implicit bias rest on their alleged general character. For instance, the conviction is that findings of studies on implicit bias can be applied to 'economically and socially important decisions, such as hiring, educational admission, and personnel evaluations' (Greenwald and Banaji, 1995: 7). Thus, a wave of organizational bias training has emerged, based on the popularized implicit bias research (MacDonald, 2017; Machery, 2022). These practices presume that research in cognitive and social psychology identifies fundamental cognitive processes or principles of decision-making shared by all people.

However, the claim about the general and generalizable character of this research is methodologically questionable. In the ongoing theoretical and methodological discussions, the external validity of many research results on heuristics-and-biases and decision-making has been called

into doubt. For instance, it has been argued that the occurrence of the phenomenon of loss aversion, which is postulated by prospect theory and widely used in practical contexts as a robust finding, is in fact dependent on the cultural factors and individual differences between the subjects studied (e.g. Apicella et al., 2014). In his later experimental work testing hypotheses derived from prospect theory, Kahneman also admits difficulties with extrapolating results from a studied group to other groups or to the whole population (Novemsky and Kahneman, 2005).

The external validity of research on implicit biases is under scrutiny as well. It has been pointed out that reliable predictions of individual behavior are hampered because the dependencies between implicit biases and behavior seem to be highly sensitive to context and unstable over time (Forscher et al., 2019; Gawronski, 2019). At a general level, it has even been argued that many results of cognitive and implicit bias research are obtained only in experimental settings, in highly controlled environments (Cesario, 2022). This argument has also been advanced in relation to the predictive power of dual process theories (Gigerenzer and Brighton, 2009; Keren, 2013).

Yet, despite the ambiguity of the concept of bias and the questionable universalism of heuristics and biases and implicit bias research, in their marketing rhetoric the AI solutions providers seem to have adopted the view of the causal efficacy of implicit biases and cognitive biases. In other words, they regard discrimination to be a result of biased human information processing. From this perspective, the challenge of discrimination is seen as solvable by replacing biased human information processing with non-biased computer processing. However, for those who consider that the phenomenon of discrimination can be better explained by structures, dynamic power relations, or performative social interactions, the idea that AI could help to reduce discrimination does not seem to be an easy and obvious solution (Hoffmann, 2019).

The arguments presented above point out that the science behind AI-based personnel selection is often misunderstood and that using research on 'irrational' and 'biased' human decision-making as a basis for anti-discrimination recruitment policies in organizations is not justified in the light of the scientific findings. But the concept of human bias is not the only ambiguous concept in the justifications for AI-based personnel selection. As we will argue in the sections below, the concepts of fairness and objectivity are likewise not as unproblematic as is often presumed in the claims supporting the use of AI in personnel selection.

Personnel selection should be fair and AI-based personnel selection guarantees fairness

The justification for the use of AI in personnel selection focuses on how individuals should be treated within the recruitment processes. The idea is that the removal of cognitive and implicit biases and replacing human decision-making with unbiased AI systems that focus on individuals' skills, merit and performance makes personnel selection fairer (claims 3 and 4). This means that the claims presume the concept of procedural fairness which focuses on how individuals are treated within the recruitment processes and which states that similar applicants should be treated equally to avoid disparate treatment (Friedler et al., 2021).

However, the procedural understanding of fairness has limitations. It might be that a recruitment procedure such as an AI-based personnel selection system is procedurally fair and treats similar applicants equally and impartially. Nevertheless, applying such a procedure might still lead to outcomes such as an allocation of jobs, salaries and wealth between demographic groups that we consider discriminatory or otherwise morally or politically unacceptable (Barocas and Selbst, 2016; Eidelson, 2021). This kind of argument has also been advanced within computer science by Jacobs and Wallach (2021) who have stressed the importance of making explicit the political and philosophical assumptions that underlie different mathematical operationalizations of fairness. One important result of the computer science debates on fairness is that the notions of procedural and outcome-based fairness[9] seem to be incompatible. Optimizing for procedural fairness does not mean achieving outcome-based fairness (Lipton et al., 2018). This contradicts the justification for relying on AI: that replacing the discriminatory decision-making of biased humans with objective and unbiased AI systems that perform equal treatment of similar individuals would solve discrimination in personnel selection.[10]

Furthermore, it should be noted that some AI vendors, such as Pymetrics and HireVue, rely on the procedural conception of fairness in marketing their systems in which they promise to remove human bias (Bogen and Rieke, 2018). However, in practice their system development is based on outcome-based notion of fairness. These systems try to optimize the selection rates between members of different demographic groups to follow the disparate impact ⅘ rule[11] of the US anti-discrimination law (Raghavan et al., 2020; Sánchez-Monedero et al., 2020). However, this outcome-based approach does not address the procedural unfairness that an AI system might produce toward the rejected candidates. For instance, if an AI system recommends hiring from two demographic groups at an equal rate (or according to the ⅘ rule), this might create accusations of procedural unfairness by members of the majority group who are qualified, but who believe that without the application of the rule they would have been hired instead of the members of minority group (Morse et al., 2022). It is an example of a dilemma concerning group quotas in affirmative action discussed in political philosophy (Arneson, 2015). We point out that no advancement in AI

optimization can provide an answer to the question of whether introducing quotas in personnel selection is right or not.

Our purpose is not to discuss whether AI development should be based on procedural or outcome-based fairness. We stress that this issue is a classical question of political philosophy that concerns all forms of personnel selection irrespective of the technology or means used to make the selection. Therefore, amidst the hype of the AI ethics focused on technical de-biasing (Hunkenschroer and Luetge, 2022; Young et al., 2022), we should also look beyond the mainstream discourse and acknowledge the non-technical, political and often conflicting nature of the fairness assumptions that the AI solutions are based on (Benbouzid, 2023; Verma and Rubin, 2018).

Personnel selection should be objective and AI-based personnel selection guarantees objectivity

As in the case of the concept of fairness, a specific notion of objectivity is presumed in the justifications for bringing AI into personnel selection (claim 4). We call this notion psychometric objectivity.[12] It refers to an aspiration to eliminate the influence of personal judgment from the standardized measurement and testing processes of individual characteristics so that the test results do not vary depending on the person conducting the measurement (Wijsen et al., 2021). Such elimination is based on attempts to control statistically for human biases to reduce the risk of discrimination (Wijsen et al., 2021). Psychometric objectivity is presumed in AI-based personnel selection since AI-based personnel selection technologies rely heavily on psychometric testing. Therefore, AI-based personnel selection can be shown to suffer from similar problems that have haunted psychometric testing in recruitment used to strive for objectivity in decisions about personnel selection.

Organizations have used psychometric measurements such as IQ and personality testing in the evaluation of employees and in the selection process for decades (Searle and Al-Sharif, 2018). These practices are based on the belief that scientifically well-validated and standardized measurement tools produce the same accurate results irrespective of who conducts the measurement (Wijsen et al., 2021). It is noteworthy that AI-based recruitment does not abandon the traditional tools of psychometric measurement. The proponents of AI-based personnel selection presume that psychometric tools are an objective and accurate way to measure candidates' suitability for a job and to predict their future job performance (Dencik and Stevens, 2021, Drage and Mackereth, 2022, Raghavan et al., 2020).

However, AI-based personnel selection suffers from problems of validation that undermine claims of objectivity. AI systems use the concepts of psychometrically well-validated personality theories, such as Big 5 personality assessment. However, they also often rely on personality typing based on methods, such as social media profiles or video interviews, that have seldom been validated (Raghavan et al., 2020; Rhea et al., 2022). For instance, an audit of two AI-based personnel selection systems showed that reliability of these systems is so low that they cannot be considered valid personality assessment instruments (Rhea et al. 2022). AI systems also face a problem with reliably predicting future employee performance based on psychometric concepts such as intelligence (Friedler et al., 2021). Predictive validity has been a challenge for psychometric recruitment testing for decades (Searle and Al-Sharif, 2018). AI systems seem to be affected by it as well. As in the case of psychometric tools, it is hard to determine the 'future performance' that AI systems should be predicting based on candidate data, such as IQ scores (Barocas and Selbst, 2016; Friedler et al., 2021).

Despite being an epistemic value, objectivity also has a political dimension.[13] This leads to the organizational political problem of recruitment that seeks psychometric objectivity: the narratives of objective AI-based personnel selection that is based on psychometric testing shifts attention away from organizational politics. Decision-making processes such as psychometric testing that are labeled as 'objective' tend to affect power relations in organizations when leaders aim to remove subjective decision-making by standardizing decision-making processes (Wijsen et al., 2021). This is also true in the case of AI-based personnel selection. AI systems in organizations can be seen as a form of algorithmic management: they are technologies which influence existing power structures between top-level leaders, mid-level managers and employees within organizations (Jarrahi et al., 2021). AI systems are used to renegotiate the meaning of 'organizational fit' by presenting soft skills such as cognitive and persona-centric evaluations as central factors for fair selection (Dencik and Stevens, 2021). The standardization of decision-making processes by AI solutions dehumanizes the hiring process (Fritts and Cabrera, 2021) and supports a power asymmetry between organizations and their employees (Yam and Skorburg 2021) in the sense that human recruiters (Li et al., 2021) and even mid-level managers become easily replaceable in the eyes of the organizations (and their top-level leaders) when the power to make decisions is, at least in part, outsourced to machine learning (Noponen, 2019). It is also noteworthy that the transfer to AI-based decision-making is usually initiated by top-level managers, and it has been shown that support for AI positively correlates with high rank in the organization (Kolbjørnsrud et al., 2017). In these developments, AI-based personnel selection is a tool of organizational politics. It emphasizes the power of the top-level leaders by aiming to replace the subjective decisions of mid-level managers with objective and fair AI systems.

In the case of AI-based personnel selection, objectivity has also been used as a label to mask societal structural inequalities and structural discrimination in the form of an apparently neutral and non-discriminative policy which has been based on psychometric testing. For instance, Drage and Mackereth (2022) discuss AI-based personnel selection systems that make use of psychometrically valid personality tests to reveal the personality of an applicant. These systems are promoted to measure personality in an objective and fair way, but Drage and Mackereth (2022) argue that they may actively reproduce both organizational and societal discrimination by masking the pre-existing structural inequalities. The authors note that the AI-based personnel evaluation systems trained on past data of successful employees reproduce social norms that constitute a good employee. The problem is not only theoretical since, in fact, many AI firms leave it up to the employer to decide on the criteria for a 'good employee' (Li et al., 2021). For instance, if an employer chooses to base hiring decisions on predicted tenure and if gender accounts for differences in tenure, this might result in discrimination against women who statistically have higher turnover (Barocas and Selbst, 2016). Such structural discrimination cannot be straightforwardly corrected by de-biasing datasets and algorithms (Barocas and Selbst, 2016). However, this problem is seldom addressed when AI solutions are promoted as more objective and fairer than human decision-making (Barocas and Selbst, 2016; Crawford 2021: 130).

Above, we presented three types of critique of the psychometric objectivity presumed in the AI-based personnel selection: problems of validity, the organizational political nature of removing subjective bias from decision-making, and the sociopolitical critique that fair and objective AI-based personnel selection masks structural discrimination. In addition, we have shown that criticism of these kinds also concerns the use of psychometric testing in recruitment. By demonstrating the similarities between the problems of using psychometric tools in traditional hiring and incorporating these tools into machine learning in recruitment, we aim to point out that the criticism of the objectivity of AI-based personnel selection is not a historically unique feature of the recent critical discussions of AI.[14] Therefore, we claim that the use of AI can be understood as the latest trend in the practices of psychometric-based recruitment.

Biases in AI-based personnel selection will be fixed through further technological development

The claim that the technical bias problems in the current AI systems can be fixed by technical means through further AI development (claim 5) is present in HRM, computer science and AI. It has been pointed out in the general criticisms of current societal uses of AI that statements like claim 5 are attempts by the AI industry to set the tone of debates about the ethics of AI: the discussion of the ethical aspects of AI has focused on what the developers are trained to do and good at doing, namely coding (Bui and Noble, 2020; Miceli et al., 2022; Slee, 2020; Young et al., 2022). In this discussion, the problems of algorithmic discrimination are also thought to have a technical fix (Bui and Noble, 2020; Crawford, 2021). However, the focus on technical bug fixing, and the talk of 'AI fairness' that goes with it, serves the strategic goals of the technology companies in protecting them from public backlash, government audits and the lawsuits that may follow cases of AI discrimination (Slee, 2020).

To complement the existing critical discussions, we wish to point out that the widespread understanding of AI as a technology of bug fixing is related to a specific presumption about the nature of AI: that human cognition can be equated with algorithmic information processing. The metaphor of 'the human mind as a computer' informed research in AI (Crawford, 2021: 4–5) and in cognitive psychology (discussed above), and became widely accepted in Western culture (Turkle, 2005). As argued by Crawford (2021: 5), the metaphor leads to viewing both the human mind and technology as detached from social, cultural and political surroundings. This idea of detachment underlies the claims that AI is an objective technological tool and that to solve societal problems such as discrimination, one only needs to tamper with algorithms and to fix datasets technically. However, such an idea of detachment is questionable. As we have argued, the current forms of AI-based personnel selection are deeply rooted in the views of organizational decision-making organizational power, as well as the fairness assumptions of modern recruitment.

The bug-fixing strategy and the idea of detachment it presumes, once again, point to interesting similarities. Let us compare the technical AI bug fixing strategies to the ways in which issues of discrimination have previously been addressed in psychometrics. In psychometrics, concerns about discriminative measurement methods have grown since the second half of the twentieth century and have led to the development of mathematical definitions of measurement bias and item bias (Wijsen et al., 2021). The mathematical solutions seem to have alleviated these concerns among proponents and users of psychometrics: '[s]tandardized testing, now including some statistics that detect possible biased items, is still considered one of psychometrics' main contributions to society' (Wijsen et al., 2021: 9). Based on this similarity, following other critical scholars (D'ignazio and Klein, 2020; Miceli et al., 2022; Young et al., 2022), we anticipate that unless the critique of the deeply political nature of the objectivity and fairness claims of the AI industry is taken seriously in policy and AI ethics, debates about AI ethics will continue serving the AI industry and will result in an image of neutrality and value

freedom similar to that which psychometrics has developed within recruitment services (and society at large).

Conclusion

We have argued that to understand the role of AI in recruitment (personnel selection) we need to unpack the claims and narratives that justify the adoption of AI in recruitment solutions. We have shown that a specific understanding of AI and its role in recruitment is shared across diverse scholarly and practical contexts – in the AI industry and development, research on cognitive and implicit biases, and in recruitment. We have also pointed out that the existing and widely accepted justifications for including AI in personnel selection make sense only if one presumes a particular understanding of the notions of rationality, bias, fairness, objectivity and AI. However, we have argued that the way in which these notions are understood is debatable from a philosophical, methodological and political point of view. What is more, the seemingly novel and disruptive character of AI-based personnel selection technologies can be challenged by showing how the ethical and political problems of AI-based recruitment are similar to the traditional problems of psychometric recruitment tools.

Our analysis refers to the critical perspectives on AI technologies from the fields of legal studies, critical management studies, computer science, sociology and science and technology studies. We believe that bringing them together allows us to identify and scrutinize, the presumptions upon which the justifications for relying on AI to address discrimination in recruitment rest. We complement the existing critical discussion with the perspective of philosophy of science. We show that the justificatory claims supporting AI-based personnel selection are statements of belief which hide their political nature behind the wordings of agreeable ethics. However, we point out that the consensus around understanding AI ethics as 'technical, apolitical, documentation-related, and unregulated' (Young et al., 2022: 1375) exceeds far beyond the AI industry and academic research and debates in computer science and AI. The justificatory claims we identified also permeate behavioral science. We hope that our article will inspire further studies of the historical and conceptual affinities between psychometric tools and AI software used in recruitment, as well as of the incorporation of research on cognitive psychology and implicit biases into the fields of recruitment and AI.

Acknowledgments

We would like to thank Caterina Marchionni, Kamil Mamak, Ilmari Hirvonen, David Moats and Tuukka Kaidesoja for their insightful comments on the previous versions of this article. We are also grateful for the audiences at the CEPDISC'22 Conference on Discrimination (Vejle, 2022), TINT Centre for Philosophy of Social Science seminar (Helsinki, 2022) and Forget Algorithmic Drama workshop (Helsinki, 2022).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study has received funding from the University of Helsinki's three-year research project 'From cyborg origins of modern economics to its automated future. Towards a new philosophy of economics', Kone Foundation (grant number 2021104708), as well as from the funds of the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 754513 and the Aarhus University Research Foundation.

ORCID iDs

Päivi Seppälä https://orcid.org/0000-0003-2429-240X
Magdalena Małecka https://orcid.org/0000-0001-5395-9256

Supplemental material

Supplemental material for this article is available online.

Notes

1. We use the term AI to refer to a wide range of technologies that are thought to 'intelligently' process information for various uses. The term is used extensively in marketing, whereas computer scientists talk mainly about specific technologies such as machine learning (Crawford, 2021). In the case of recruitment, AI technologies are most often machine learning algorithms, natural language processing and computer vision (Sloane et al., 2022). In recruitment, the task is to find patterns in sets of learning data to make predictions about job applicant's future job performance (Li et al., 2021).
2. We characterize discrimination as disadvantageous differential treatment that is based on perceived membership in a socially salient group (Lippert-Rasmussen, 2013).
3. We use the term recruitment to refer to 1) the applicant attraction phase and 2) the applicant evaluation and selection phase of the recruitment process (Searle and Al-Sharif, 2018). We use the term personnel selection to refer to the evaluation of the candidates and making hiring decisions.
4. Often discussed examples of algorithmic bias are Amazon's AI-based personnel selection system that discriminated against women (Dastin, 2018) and COMPAS software which is used in the US justice system to predict the likelihood of a person committing a crime and which has been accused of being biased against African Americans (Angwin et al., 2016). See O'Neill (2016) for an overview.
5. See also Lin et al. (2021) and Soleimani et al. (2021).
6. We do not suggest that every person working in the field of AI marketing, computer science, AI ethics, in behavioural science or in HR studies endorses these views. For instance, we acknowledge that some authors stress that technological

solutions to biases need to be complemented by incorporating organizational decision-makers into AI development to overcome AI-based discrimination in recruitment (Bogen and Rieke, 2018; Cho et al., 2023; Köchling and Wehner, 2020; Nadeem et al., 2021; Vassilopoulou et al., 2022).

7. It is not obvious how the alleged irrationality of human decision-making in the form of cognitive biases is linked to discrimination, even though Kahneman et al. (2021: 335) suggest this by claiming that algorithms reduce discrimination when they replace human cognitive biases. Only in the theories of implicit bias, which attempt to causally explain discrimination, is the connection clear: implicit attitudes and stereotypes about socially salient groups may affect decision-making about these groups (Greenwald and Banaji, 1995).
8. Implicit bias theories are often interpreted along the lines of some dual-process theory (Payne and Gawronski, 2010).
9. Outcome-based fairness concerns how outcomes (for instance, the allocation of jobs) are distributed between members of social groups (Friedler et al., 2021).
10. We thank one of the reviewers for this point.
11. The selection rate of candidates of the highest-passing and lowest-passing demographic groups should not be smaller than 0.80, otherwise the selection method might be considered discriminatory (Sánchez-Monedero et al., 2020).
12. Compare with Green and Viljoen (2020) who call this understanding of objectivity and neutrality of AI systems 'algorithmic formalism'.
13. Compare with points raised in the context of AI debates by Mark Coeckelbergh (2022) on justice and democracy.
14. The idea of algorithms as a contested way to achieve objectivity and societal fairness reaches beyond the introduction of modern psychometric testing. Ochigame (2020) traces the roots of the idea of 'algorithmic fairness' in the 17th century accounting practices and argues that algorithmic fairness should be understood as an aspiration that reappears persistently in history.

References

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias. *Pro Publica* 23(May).

Apicella CL, Azevedo EM, Christakis NA, et al. (2014) Evolutionary origins of the endowment effect: Evidence from hunter-gatherers. *American Economic Review* 104(6): 1793–1805.

Arneson R (2015) Equality of opportunity. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy* (Summer 2015 ed). Stanford, CA: The Metaphysics Research Lab. Available at: https://plato.stanford.edu/archives/sum2015/entries/equal-opportunity/ (accessed 17 July 2023).

Bailao Goncalves M, Anastasiadou M and Santos V (2022) AI and public contests: A model to improve the evaluation and selection of public contest candidates in the police force. *Transforming Government: People, Process and Policy* 16(4): 627–648.

Barocas S and Selbst AD (2016) Big data's disparate impact. *California Law Review* 104(3): 671–732.

Benbouzid B (2023) Fairness in machine learning from the perspective of sociology of statistics: How machine learning is becoming scientific by turning its back on metrological realism. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, USA, June 12–15, 2023, pp. 35–43. New York: ACM.

Birzhandi P and Cho Y-S (2023) Application of fairness to healthcare, organizational justice, and finance: A survey. *Expert Systems with Applications* 216: e119465.

Bogen M and Rieke A (2018) Help wanted - An examination of hiring algorithms, equity, and bias. Report, Upturn, December.

Bui ML and Noble SU (2020) We're missing a moral framework of justice in artificial intelligence. In: Dubber MD, Pasquale F and Das S (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 163–179.

Burrell J and Fourcade M (2021) The society of algorithms. *Annual Review of Sociology* 47(1): 1–25.

Canditech (2022) Leveraged technology to reach your hiring goals. Available at: https://www.canditech.io/technology (accessed 8 June 2022).

Cesario J (2022) What can experimental studies of bias tell us about real-world group disparities? *Behavioral and Brain Sciences* 45: E66.

Chilunjika A, Intauno K and Chilunjika SR (2021) Artificial intelligence and public sector human resource management in South Africa: Opportunities, challenges and prospects. *SA Journal of Human Resource Management* 20(1): 1–12.

Cho W, Choi S and Choi H (2023) Human resources analytics for public personnel management: Concepts, cases, and caveats. *Administrative Sciences* 13(2): 1–22.

Coeckelbergh M (2022) *The Political Philosophy of AI: An Introduction*. Cambridge: John Wiley & Sons.

Crawford K (2021) *The Atlas of AI*. New Haven and London: Yale University Press.

Dastin J (2018) Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 11 October. Available at: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (accessed 3 March 2023).

Dencik L and Stevens S (2021) Regimes of justification in the datafied workplace: The case of hiring. *New Media & Society*: 1–19.

Dennis MJ and Aizenberg E (2022) The ethics of AI in human resources. *Ethics and Information Technology* 24: e25.

D'ignazio C and Klein LF (2020) *Data feminism*. Cambridge, MA: MIT press.

Drage E and Mackereth K (2022) Does AI debias recruitment? Race, gender, and AI's "eradication of difference". *Philosophy and Technology* 35: e89.

Edwards PN (1996) *The Closed World: Computers and the Politics of Discourse in Cold War America*. Cambridge, MA: MIT Press.

Eidelson B (2021) Patterned inequality, compounding injustice, and algorithmic prediction. *American Journal of Law and Equality* 1: 252–276.

Erickson P, Klein J, Daston L, et al. (2013) *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. Chicago: University of Chicago Press.

Eronen MI (2020) Causal discovery and the problem of psychological interventions. *New Ideas in Psychology* 59: e100785.

Fernandes França TJF, Mamede HS, Perreira Barroso JM, et al. (2023) Artificial intelligence applied to potential assessment

and talent identification in an organisational context. *Heliyon* 9(4): e14694.

Forscher PS, Lai CK, Axt JR, et al. (2019) A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology* 117(3): 522–559.

Friedler SA, Scheidegger C and Venkatasubramanian S (2021) The (im)possibility of fairness. *Communications of the ACM* 64(4): 136–143.

Fritts M and Cabrera F (2021) AI Recruitment algorithms and the dehumanization problem. *Ethics and Information Technology* 23: 791–801.

Gawronski B (2019) Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Sciences* 14(4): 574–595.

Gigerenzer G (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review* 103(3): 592–596.

Gigerenzer G and Brighton H (2009) Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science* 1(1): 107–143.

Grant MJ and Booth A (2009) A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal* 26(2): 91–108.

Grayot JD (2020) Dual process theories in behavioral economics and neuroeconomics: A critical review. *Review of Philosophy and Psychology* 11: 105–136.

Green B and Viljoen S (2020) Algorithmic realism: expanding the boundaries of algorithmic thought. In Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January, 2020, pp. 19–31. New York: ACM.

Greenwald AG and Banaji MR (1995) Implicit social cognition - attitudes, self-esteem, and stereotypes. *Psychological Review* 102(1): 4–27.

Greenwald AG, Dasgupta N, Dovidio JF, et al. (2022) Implicit-Bias remedies: Treating discriminatory bias as a public-health problem. *Psychological Science in the Public Interest* 23(1): 7–40.

Greenwald AG, McGhee DE and Schwartz JLK (1998) Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74(6): 1464–1480.

HireVue (2022) Increase diversity and mitigate bias. Available at: https://www.hirevue.com/employment-diversity-bias (accessed 8 June 2022).

Hofeditz L, Mirbabaie M, Luther A, et al. (2022) Ethics Guidelines for Using AI-based Algorithms in Recruiting: Learnings from a Systematic Literature Review. In: Proceedings of the 55th Hawaii International Conference on System Sciences (ed Bui T), Manoa, Hawaii, January 3–7, 2022, pp. 145–154. Manoa: University of Hawaii.

Hoffmann AL (2019) Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22(7): 900–915.

Houwer JD, Gawronski B and Barnes-Holmes D (2013) A functional-cognitive framework for attitude research. *European Review of Social Psychology* 24(1): 252–287.

Hunkenschroer AL and Luetge C (2022) Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics* 178(4): 977–1007.

Jacobs AZ and Wallach H (2021) Measurement and fairness. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event, Canada, March 3–10, 2021, pp. 375–385. New York: ACM.

Jarrahi MH, Newlands G, Lee MK, et al. (2021) Algorithmic management in a work context. *Big Data and Society* 8(2): 20539517211020332.

Kahneman D (2011) *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman D, Sibony O and Sunstein CR (2021) *Noise: A Flaw in Human Judgment*. New York: Hachette.

Kahneman D, Slovic SP, Slovic P and Tversky A (eds) (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman D and Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* 47(2): 263–291.

Kaur G and Kaur R (2022) A critical review on analysis of human resource functions using AI technologies. *AIP Conference Proceedings* 2555(1): e020004.

Keren G (2013) A tale of two systems: A scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich. *Perspectives on Psychological Science* 8(3): 257–262.

Kleinberg J, Ludwig J, Mullainathan S, et al. (2018) Discrimination in the age of algorithms. *Journal of Legal Analysis* 10: 113–174.

Köchling A and Wehner MC (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* 13(3): 795–848.

Kolbjørnsrud V, Amico R and Thomas RJ (2017) Partnering with AI: How organizations can win over skeptical managers. *Strategy & Leadership* 45: 37–43.

Li L, Lassiter T, Oh J, et al. (2021) Algorithmic hiring in practice: Recruiter and HR professional's perspectives on AI use in hiring. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual Conference, May 19–21, 2021, pp. 166–176. New York: ACM.

Lin YT, Hung TW and Huang LTL (2021) Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy & Technology* 34: 65–90.

Lippert-Rasmussen K (2013) *Born Free and Equal?: A Philosophical Inquiry into the Nature of Discrimination*. Oxford: Oxford University Press.

Lipton Z, McAuley J, Chouldechova A (2018) Does mitigating ML's impact disparity require treatment disparity? In: *Advances in Neural Information Processing Systems 31* (eds Bengio S et al.), Montreal, Canada. 3–8 December 2018, pp. 8125–8135. New York: Curran Associates, Inc.

MacDonald H (2017) The false "science" of implicit bias. *Wall Street Journal* 9(October).

Machery E (2022) Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews: Cognitive Science* 13(1): e1569.

Małecka M (2021) Knowledge, behaviour, and policy: Questioning the epistemic presuppositions of applying behavioural science in public policymaking. *Synthese* 199: 5311–5338.

Marr D and Poggio T (1976) From understanding computation to understanding neural circuitry. Report, MIT Artificial intelligence laboratory, AI Memo 357.

Miceli M, Posada J and Yang T (2022) Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction* 14(6): 1–4.

Miller GA (2003) The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences* 7(3): 141–144.

Morse L, Teodorescu MHM, Awwad Y, et al. (2022) Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics* 181(1): 1083–1095.

Nadeem A, Marjanovic O and Abedin B (2021) Gender bias in AI: Implications for managerial practices. In: Dennehy D (eds) *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*. Cham: Springer, 259–270.

Noponen N (2019) Impact of artificial intelligence on management. *Electronic Journal of Business Ethics and Organization Studies* 24(2): 43–50.

Novemsky N and Kahneman D (2005) The boundaries of loss aversion. *Journal of Marketing Research* 42(2): 119–128.

Ochigame R (2020) The Long History of Algorithmic Fairness. Phenomenal World January 30. Available at: https://www.phenomenalworld.org/analysis/long-history-algorithmic-fairness.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Payne BK and Gawronski B (2010) A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In: Gawronski B and Payne BK (eds) *Handbook of Implicit Social Cognition: Measurement, Theory and Applications*. New York: Guildford Press, 2010: 1–15.

Pessach D and Shmueli E (2022) A review on fairness in machine learning. *ACM Computing Surveys* 55(3): 1–44.

Raghavan M, Barocas S, Kleinberg J, et al. (2020) Mitigating bias in algorithmic hiring. In: Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January, 2020, pp. 469–481. New York: ACM.

Rhea AK, Markey K, D'Arinzo L, et al. (2022) An external stability audit framework to test the validity of personality prediction in AI hiring. *Data Mining and Knowledge Discovery* 36(1): 2153–2193.

Sánchez-Monedero J, Dencik L and Edwards L (2020) What does it mean to "solve" the problem of discrimination in hiring? In: Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January, 2020, pp. 458–468. New York: ACM.

Searle RH and Al-Sharif R (2018) Recruitment and selection. In: Collings DG, Wood GT and Szamosi LT (eds) *Human Resource Management - A Critical Approach*. London: Routledge, 215–237.

Simon HA (1984) *Models of Bounded Rationality, Volume 1: Economic Analysis and Public Policy*. MIT Press Books.

Slee T (2020) The incompatible incentives of private-sector AI. In: Dubber MD, Pasquale F and Das S (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 109–123.

Sloane M, Moss E and Chowdhury RA (2022) Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns* 3(2): 100425.

Soleimani M, Intezari A, Taskin N, et al. (2021) Cognitive biases in developing biased artificial intelligence recruitment system. In: Proceedings of the 54th Hawaii International Conference on System Sciences, (ed Bui T), Manoa, Hawaii, January 4–8, 2021, pp. 5091–5099. Manoa: University of Hawaii.

Storm KIL, Reiss LK, Günther E, et al. (2023) Unconscious bias in the HRM literature: Towards a critical-reflexive approach. *Human Resources Management Review* 33(3): e100969.

Sunstein C (2019) Algorithms, correcting biases. *Social Research: An International Quarterly* 86(2): 499–511.

Thaler RH (2000) From *Homo economicus* to *Homo sapiens*. *The Journal of Economic Perspectives* 14(1): 133–141.

Thaler RH and Sunstein CR (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*. London: Penguin.

Turkle S (2005) *The Second Self: Computers and the Human Spirit*. Cambridge, MA: MIT Press.

Tursunbayeva A, Pagliari C, Lauro SD, et al. (2022) The ethics of people analytics: Risks, opportunities and recommendations. *Personnel Review* 51(3): 900–921.

Tversky A and Kahneman D (1974) Judgment under uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185(4157): 1124–1131.

Vassilopoulou J, Kyriakidou O, Özbilgin MF, et al. (2022) Scientism as illusio in HR algorithms: Towards a framework for algorithmic hygiene for bias proofing. *Human Resources Management Journal* 2022(1): 1–15.

Verma S and Rubin J (2018) Fairness definitions explained. In: Proceedings of the International Workshop on Software Fairness, Gothenburg, Sweden, May 29, 2018 pp. 1–7. New York: ACM.

Wijsen LD, Borsboom D and Alexandrova A (2021) Values in psychometrics. *Perspectives on Psychological Science* 17(3): 788–804.

Will P, Krpan D and Lordan G (2023) People versus machines: Introducing the HIRE framework. *Artificial Intelligence Review* 56(2): 1071–1100.

Yam J and Skorburg JA (2021) From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology* 23(4): 611–623.

Young M, Katell M and Krafft PM (2022) Confronting Power and Corporate Capture at the FAccT Conference. In: Proceedings of 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–42 June, 2022, pp. 1375–1386. New York: ACM.