

Köchling, Alina; Wehner, Marius Claus

Article

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

Provided in Cooperation with:

VHB - Verband der Hochschullehrer für Betriebswirtschaft, German Academic Association of Business Research

Suggested Citation: Köchling, Alina; Wehner, Marius Claus (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, ISSN 2198-2627, Springer, Heidelberg, Vol. 13, Iss. 3, pp. 795-848, <https://doi.org/10.1007/s40685-020-00134-w>

This Version is available at:

<https://hdl.handle.net/10419/233200>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development

Alina Köchling¹  · Marius Claus Wehner¹ 

Received: 15 October 2019 / Accepted: 1 November 2020 / Published online: 20 November 2020
© The Author(s) 2020

Abstract Algorithmic decision-making is becoming increasingly common as a new source of advice in HR recruitment and HR development. While firms implement algorithmic decision-making to save costs as well as increase efficiency and objectivity, algorithmic decision-making might also lead to the unfair treatment of certain groups of people, implicit discrimination, and perceived unfairness. Current knowledge about the threats of unfairness and (implicit) discrimination by algorithmic decision-making is mostly unexplored in the human resource management context. Our goal is to clarify the current state of research related to HR recruitment and HR development, identify research gaps, and provide crucial future research directions. Based on a systematic review of 36 journal articles from 2014 to 2020, we present some applications of algorithmic decision-making and evaluate the possible pitfalls in these two essential HR functions. In doing this, we inform researchers and practitioners, offer important theoretical and practical implications, and suggest fruitful avenues for future research.

Keywords Fairness · Discrimination · Perceived fairness · Ethics · Algorithmic decision-making in HRM · Literature review

1 Introduction

Algorithmic decision-making in human resource management (HRM) is becoming increasingly common as a new source of information and advice, and it will gain more importance due to the rapid growth of digitalization in organizations.

✉ Alina Köchling
alina.koechling@hhu.de

¹ Faculty of Business Administration and Economics, Heinrich-Heine-University Düsseldorf, Universitätsstrasse 1, 40225 Dusseldorf, Germany

Algorithmic decision-making is defined as automated decision-making and remote control, as well as standardization of routinized workplace decisions (Möhlmann and Zalmanson 2017). Algorithms, instead of humans, make decisions, and this has important individual and societal implications in organizational optimization (Chalfin et al. 2016; Lee 2018; Lindebaum et al. 2019). These changes in favor of algorithmic decision-making make it easier to discover hidden talented employees in organizations and review a large number of applications automatically (Silverman and Waller 2015; Carey and Smith 2016; Savage and Bales 2017). In a survey of 200 artificial intelligence (AI) specialists from German companies, 79% stated that AI is irreplaceable for competitive advantages (Deloitte 2020). Several commercial providers, such as Google, IBM, SAP, and Microsoft, already offer algorithmic platforms and systems that facilitate current human resource (HR) practices, such as hiring and performance measurements (Walker 2012). In turn, well-known and large companies, such as Vodafone, Intel, Unilever, and Ikea, apply algorithmic decision-making in HR recruitment and HR development (Daugherty and Wilson 2018; Precire 2020).

The major driving forces for algorithmic decision-making are savings in both costs and time, minimizing risks, enhancing productivity, and increasing certainty in decision-making (Suen et al. 2019; McDonald et al. 2017; McColl and Michelotti 2019; Woods et al. 2020). Besides these economic reasons, firms seek to diminish the human biases (e.g., prejudices and personal beliefs) by applying algorithmic decision-making, thereby increasing the objectivity, consistency, and fairness of the HR recruitment as well as HR development processes (Langer et al. 2019; Florentine 2016; Raghavan et al. 2020). For example, Deloitte argues that the algorithmic decision-making system always manages each application with the same attention according to the same requirements and criteria (Deloitte 2018). At first glance, algorithmic decision-making seems to be more objective and fairer than human decision-making (Lepri et al. 2018).

However, there is a possible threat of discrimination and unfairness by relying solely on algorithmic decision-making (e.g., (Lee 2018; Lindebaum et al. 2019; Simbeck 2019)). In general, discrimination is defined as the unequal treatment of different groups based on gender, age, or ethnicity instead of on qualitative differences, such as individual performance (Arrow 1973). Algorithms produce discrimination or biased outcomes if they are trained on inaccurate (Kim 2016), biased (Barocas and Selbst 2016), or unrepresentative input data (Suresh and Gutttag 2019). Consequently, algorithms are vulnerable to produce or replicate biased decisions if their input (or training) data are biased (Chander 2016).

Complicating this issue, biases and discrimination are often only recognized after algorithms have made a decision. As a prominent example stemming from the current debate around transparency, bias, and fairness in algorithmic decision-making (Dwork et al. 2012; Lepri et al. 2018; Diakopoulos 2015), the hiring algorithms applied by the American e-commerce specialist Amazon yielded an extreme disadvantage of female applicants, which finally led Amazon to shut down the complete algorithmic decision-making for their hiring decision (Dastin 2018; Miller 2015). Thus, the lack of transparency and accountability of the input data, the algorithm itself, and the factors influencing algorithmic outcomes are potential

issues associated with algorithmic decision-making (Citron and Pasquale 2014; Pasquale 2015). Another question remains whether applicants and/or employees perceive the algorithmic decision-making to be fair. Previous studies showed that applicants' and employees' acceptance of algorithmic decision-making is lower in HR recruitment and HR development compared to common procedures conducted by humans (Kaibel et al. 2019; Langer et al. 2019; Lee 2018).

Consequently, there is a discrepancy between the enthusiasm about algorithmic decision-making as a panacea for inefficiencies and labor shortages on one hand and the threat of discrimination and unfairness of algorithmic decision-making on the other side. While the literature in the field of computer science has already addressed the issues of biases, knowledge about the potential downsides of algorithmic decision-making is still in its infancy in the field of HRM despite its importance due to increased digitization and automation in HRM. This heterogeneous state of research on discrimination and fairness raises distinct challenges for future research. From a practical point of view, it is problematic if large and well-known companies implement algorithms without being aware of the possible pitfalls and negative consequences. Thus, to move the field forward, it is paramount to systematically review and synthesize existing knowledge about biases and discrimination in algorithmic decision-making and to offer new research avenues.

The aim of this study is threefold. First, this review creates an awareness of potential biases and discrimination resulting from algorithmic decision-making in the context of HR recruitment and HR development. Second, this study contributes to the current literature by informing both researchers and practitioners about the potential dangers of algorithmic decision-making in the HRM context. Finally, we guide future research directions with an understanding of existing knowledge and gaps in the literature. To this end, the present paper conducts a systematic review of the current literature with a focus on HR recruitment and HR development. These two HR functions deal with the potential of future and current employees and the (automatic) prediction of person-organization fit, career development, and future performance (Huselid 1995; Walker 2012). Decisions made by algorithms and AI in these two important HR areas have serious consequences for individuals, the company, and society concerning ethics and both procedural and distributive fairness (Ötting and Maier 2018; Lee 2018; Tambe et al. 2019; Cappelli et al. 2020).

Our study contributes to the existing body of research in several ways. First, the systematic literature review contributes to the literature by highlighting the current debate on ethical issues associated with algorithmic decision-making, including bias and discrimination (Barocas and Selbst 2016). Second, our research provides illustrative examples of various algorithmic decision-making tools used in HR recruitment, HR development, and their potential for discrimination and perceived fairness. Moreover, our systematic review underlines the fact that it is a timely topic gaining enormous importance. Companies will face legal and reputational risk if their HR recruitment and HR development methods turn out to be discriminatory, and applicants and employees may consider the algorithmic selection or development process to be unfair.

For this reason, companies need to know that the use of algorithmic decision-making can yield to discrimination, unfairness, and dissatisfaction in the context of

HRM. We offer an understanding of how discrimination might arise when implementing algorithmic decision-making. We try to give guidance on how discrimination and perceived unfairness could be avoided and provide detailed directions for future research in the existing literature, especially in the HRM field. Moreover, we identify several research gaps, mainly a lacking focus on perceived fairness.

The paper is organized as follows: first, we give an understanding of key terms and definitions. Afterward, we present the methodology of our systematic literature review accompanied by a descriptive analysis of the reviewed literature. This is followed by an illustration of the current state of knowledge on algorithmic decision-making and subsequent discussion. Finally, we offer practical as well as theoretical implications and outline future research avenues.

2 Conceptual background and definitions

2.1 Definition of algorithms

The Oxford Living Dictionary defines algorithms as “processes or sets of rules to be followed in calculations or other problem-solving operations, especially by a computer.” Möhlmann and Zalmanson (2017) refer to algorithmic decision-making as automated decision-making and remote control, and standardization of routinized workplace decision. Thus, in this paper, we use the term algorithmic decision-making to describe a computational mechanism that autonomously makes decisions based on rules and statistical models without explicit human interference (Lee 2018). Algorithms are the basis for several AI decision tools.

AI is an umbrella term for a wide array of models, methods, and prescriptions used to simulate human intelligence, often when it comes to collecting, processing, and acting on data. AI applications can apply rules, learn over time through the acquisition of new data and information, and adapt to changes in the environment (Russell and Norvig 2016). AI includes several different research areas, such as machine learning (ML), speech and image recognition, and natural language processing (NLP) (Kaplan and Haenlein 2019; Paschen et al. 2020).

As mentioned, the basis for many AI decision-making tools used in HR are ML algorithms, which can be categorized into three major types: supervised, unsupervised, and reinforcement learning (Lee and Shin 2020). Supervised ML algorithms aim to make predictions (often divided into classification- or regression-type problems), given the input data and desired outputs considered as the ground truth. Human experts often provide these labels and thus provide the algorithm with the ground truth. To replicate human decisions or to make predictions, the algorithm learns patterns from the labeled data and develops rules, which can be applied for future instances for the same problem (Canhoto and Clear 2020). In contrast, in unsupervised ML, only input data are given, and the model learns patterns from the data without a priori labeling (Murphy 2012). Unsupervised ML algorithms capture the structural behaviors of variables in the input data for theme analysis or grouping

data (Canhoto and Clear 2020). Finally, reinforcement learning, as a separate group of methods, is not based on fixed input/output data. Instead, the ML algorithm learns behavior through trial-and-error interactions with a dynamic environment (Kaelbling et al. 1996).

Furthermore, instead of grouping ML models as supervised, unsupervised, or reinforcement type learning, the methodologies of algorithms may also be used to categorize ML models. Examples are probabilistic models, which may be used in supervised or unsupervised settings (Murphy 2012), or deep learning models (Lee and Shin 2020), which rely on artificial neural networks and perform complex learning tasks. In supervised settings, neural network models often determine the relationship between input and output using network structures containing the so-called hidden layers, meaning phases of transformation of the input data. Single nodes of these layers (neurons) were first modeled after neurons in the human brain, and they resemble human thinking (Bengio et al. 2017). In other settings, deep learning may be used, for instance, to (1) process information through multiple stages of nonlinear transformation; or (2) determine features, representations of the data providing an advantage for, e.g., prediction tasks (Deng and Yu 2014).

2.2 Reason for biases

For any estimation \hat{Y} of a random variable Y , bias refers to the difference between the expected values of \hat{Y} and Y and is also referred to as systematic error (Kauermann and Kuechenhoff 2010; Goodfellow et al. 2016). Cognitive biases, specifically, are systematic errors in human judgment when dealing with uncertainty (Kahneman et al. 1982). These cognitive biases are thought to be transferred to algorithmic evaluations or predictions, where bias may refer to “computer systems that systematically and unfairly discriminate against certain individuals or groups in favor of others” (Friedman and Nissenbaum 1996, p. 332).

Algorithms are often characterized as “black box”. In the context of HRM, Cheng and Hackett (2019) characterize algorithms as “glass boxes”, since some, but not all, components of the theory are reflective. In this context, the consideration and distinction of the three core elements are necessary, namely, transparency, interpretability, and explainability (Roscher et al. 2020). Transparency is concerned with the ML approach, while interpretability is concerned with the ML model in combination with the data, which means the making sense of the obtained ML model (Roscher et al. 2020). Finally, explainability comprises the model, the data, and human involvement (Roscher et al. 2020). Concerning the former, transparency can be distinguished at three different levels: “[...] at the level of the entire model (simulatability), at the level of individual components, such as parameters (decomposability), and at the level of the training (algorithmic transparency)” (Roscher et al. 2020, p. 4). Interpretability concerns the characteristics of an ML model that need to be understood by a human (Roscher et al. 2020). Finally, the element of explainability is paramount in HRM. Contextual information of human and their knowledge from the domain of HRM are necessary to explain the different sets of interpretations and derive conclusions about the results of the algorithms

(Roscher et al. 2020). Especially in HRM, in which ML algorithms are increasingly used for prediction of variables of interest to the HR department (e.g., personality characteristics, employee satisfaction, and turnover intentions), it is essential to understand how the ML algorithm operates (e.g., how the ML algorithm uses data and weighs specific criteria) and the underlying reasons for the produced decision.

In the following, we will outline the main reasons for biases in algorithmic decision-making and briefly summarize different biases, namely historical, representation, technical, and emergent bias. One of the main reasons for bias in algorithmic decision-making is the quality of input data, because algorithms learn from historical data as an example; thus, the learning process depends on the exposed examples (Friedman and Nissenbaum 1996; Barocas and Selbst 2016; Danks and London 2017). The input data are usually historical. Consequently, if the input data set is biased in one way or another, the subsequent analysis is biased, as well (keyword: “garbage in, garbage out”). For example, if the input data of an algorithm include implicit or explicit human judgments, stereotypes, or biases, an accurate algorithmic output will inevitably entail these human judgments, stereotypes, and prejudices (Diakopoulos 2015; Suresh and Guttag 2019; Barfield and Pagallo 2018). This bias usually exists before the creation of the system and may not be apparent at first glance. In turn, the algorithm replicates these preexisting biases, because it treats all information, in which a certain kind of discrimination or bias is embedded, as a valid example (Barocas and Selbst 2016; Lindebaum et al. 2019). In the worst case, the algorithm can yield racist or discriminatory outputs (Veale and Binns 2017). Algorithms exhibit these tendencies, even if it is not the intention of the manual programming since they compound the historical biases of the past. Thus, any predictive algorithmic decision-making tool built on historical data may inherit historical biases (Datta et al. 2015).

As an example from the recruitment process, if an algorithm is trained on historical employment data, integrating an implicit bias that favors white men over Hispanics, then, without even being fed data on gender or ethnicity, an algorithm may recognize patterns in the data, which expose an applicant as a member of a certain protected group, which, historically, is less likely to be chosen for a job interview. This, in turn, may lead to a systematic disadvantage of certain groups, even if the designer has no intention of marginalizing people based on these categories and if the algorithm is not directly given this information (Barocas and Selbst 2016).

Another reason for biases in algorithms related to the input data is that certain groups or characteristics are mostly underrepresented or sometimes overrepresented, which is also called representation bias (Barocas and Selbst 2016; Suresh and Guttag 2019; Barfield and Pagallo 2018). Any decision based on this kind of biased data might lead to disadvantages of groups of individuals who are underrepresented or overrepresented (Barocas and Selbst 2016). Another reason for representation bias can be the absence of specific information (Barfield and Pagallo 2018). Thus, not only the selection of measurements but also the preprocessing of the measurement data might yield to bias. ML models often evolve in several steps of feature engineering or model testing, since there is no universally best model (as shown in the “no free lunch” theorems, [see Wolpert and

Macready (1997)]. Here, the choice of the benchmark or rather the value indicating the performance of the model is optimized through rotations of different representations of the data and methods for prediction. For example, representative bias might occur if females in comparison to males are underrepresented in the training data of an algorithm. Hence, the outcome could be in favor of the overrepresented group (i.e., males) and, hence, lead to discriminatory outcomes.

Technical bias may arise from technical constraints or technical consideration for several reasons. For example, technical bias can originate from limited “[...] computer technology, including hardware, software, and peripherals” (Friedman and Nissenbaum 1996, p. 334). Another reason could be a decontextualized algorithm that does not manage to treat all groups fairly under all important conditions (Friedman and Nissenbaum 1996; Bozdog 2013). The formalization of human constructs to computers can be another problem leading to technical bias. Human constructs, such as judgments or intuitions, are often hard to quantify, which makes it difficult or even impossible to translate them to the computer (Friedman and Nissenbaum 1996). As an example, the human interpretation of law can be ambiguous and highly dependent on the specific context, making it difficult for an algorithmic system to correctly advise in litigation (c.f., Friedman and Nissenbaum 1996).

In the context of real users, emergent bias may arise. Typically, this bias occurs after the construction as a result of changed societal knowledge, population, or cultural values (Friedman and Nissenbaum 1996). Consequently, a shift in the context of use might yield to problems and an emergent bias due to two reasons, namely “new societal knowledge” and “mismatch between users and system design” (see Table 1 in Friedman and Nissenbaum 1996, p. 335). If it is not possible to incorporate new knowledge in society into the system design, emergent bias due to new societal knowledge occurs. The mismatch between users and system design can occur due to changes in state-of-the-art-research or due to different values. Also, emergent bias can occur if a population uses the system with different values than those assumed in the design process (Friedman and Nissenbaum 1996). Problems occur, for example, when users originate from a cultural context that avoids competition and promotes cooperative efforts, while the algorithm is trained to reward individualistic and competitive behavior (Friedman and Nissenbaum 1996).

2.3 Fairness and discrimination in information systems

Leventhal (1980) describes fairness as equal treatment based on people’s performance and needs. Table 1 offers an overview of the different fairness definitions. Individual fairness means that, independent of group membership, two individuals who are perceived to be similar by the measures at hand should also be treated similarly (Dwork et al. 2012). Rising from the micro-level onto the meso-level, Dwork et al. (2012) also proposed another measure of fairness, that is, group fairness, in which entire (protected) groups of people are required to be treated similarly (statistical parity). Hardt et al. (2016) extended these notions by including true outcomes of predicted variables to achieve fair treatment. In their sense, false-

Table 1 Definitions of fairness

Name	Author	Definition
Individual fairness	Dwork et al. (2012)	“Similar” subjects should have “similar” classifications
Group fairness		Subjects in protected and unprotected groups have an equal probability of being assigned positive $\mathbb{P}(\widehat{Y} = 1 G = 1) = \mathbb{P}(\widehat{Y} = 1 G = 0)$
Equal opportunity	Hardt et al. (2016)	False-negative rates should be equal $\mathbb{P}(\widehat{Y} = 0 Y = 1, G = 1) = \mathbb{P}(\widehat{Y} = 0 Y = 1, G = 0)$

$Y \in \{0, 1\}$ is a random variable describing, e.g., the recidivism of a subject, \widehat{Y} its estimator and $G \in \{0, 1\}$, describes whether a subject is a member of a certain protected group ($G = 1$) or not ($G = 0$)

positives/negatives are sources of disadvantage and should be equal among groups means equal opportunity for false-positives/negatives (Hardt et al. 2016).

Unfair treatment of certain groups of people or individual subjects yields to discrimination. Discrimination is defined as the unequal treatment of different groups (Arrow 1973). Discrimination is very similar to unfairness. Discriminatory categories can be strongly correlated with non-discriminatory categories, such as age (i.e., discriminatory) and years of working experience (non-discriminatory) (Persson 2016). Also, there is a difference between implicit and explicit discrimination. Implicit discrimination is based on implicit attitudes or stereotypes and often unintentional (Bertrand et al. 2005). In contrast, explicit discrimination is a conscious process due to an aversion to certain groups of people. In HR recruitment and HR development, discrimination means the not-hiring or support of a person due to characteristics not related to that person’s productivity in the current position (Frijters 1998).

The HR literature, especially the literature on personnel selection, is concerned with fairness in hiring decisions, because every selection measure of individual differences is inevitably discriminatory (Cascio and Aguinis 2013). However, the question arises “whether the measure discriminates unfairly” (Cascio and Aguinis 2013, p. 183). Hence, the actual fairness of prediction systems needs to be tested based on probabilities and estimates, which we refer to as objective fairness. In the selection context, the literature distinguishes between differential validity (i.e., differences in subgroup validity) and differential prediction (i.e., differences in slopes and intercepts of subgroups), and both might lead to biased results (Meade and Fetzer 2009; Roth et al. 2017; Bobko and Bartlett 1978).

In HR recruitment and HR development, both objective fairness and subjective fairness perceptions of applicants and employees about the usage of algorithmic decision-making need to be considered. In this regard, perceived fairness or justice is more a subjective and descriptive personal evaluation rather than an objective reality (Cropanzano et al. 2007). Subjective fairness plays an essential role in the relationship between humans and their employers. Previous studies showed that the

likelihood of conscientious behavior and altruisms is higher for employees who feel treated fairly (Cohen-Charash and Spector 2001). Conversely, unfairness can have considerable adverse consequences. For example, in the recruitment context, fairness perceptions of candidates during the selection process have important consequences for decision to stay in the applicant pool or accept a job offer (Bauer et al. 2001). Therefore, it is crucial to know how people feel about algorithmic decision-making taking over managerial decisions formerly made by humans, since the fairness perceptions during the recruitment process and/or training process have essential and meaningful effects on attitudes, performance, morale, intentions, and behavior (e.g., the acceptance or rejection of a job offer or job turnover, job dissatisfaction, and reduction or elimination of conflicts) (Gilliland 1993; McCarthy et al. 2017; Hausknecht et al. 2004; Cropanzano et al. 2007; Cohen-Charash and Spector 2001). Moreover, negative experiences might damage the employer's image. Several online platforms offer the possibility of rating companies and their recruitment and development process (Van Hove 2013; Woods et al. 2020).

Considering justice and fairness in the organizational context (Gilliland 1993), there are three core dimensions of justice: distributive, procedural, and interactional. The three dimensions tend to be correlated. Distributive justice deals with the outcome that some humans receive and some do not (Cropanzano et al. 2007). Rules that can lead to distributive justice are “[...] equality (to each the same), equity (to each in accordance with contributions, and need (to each in accordance with the most urgency)” (Cropanzano et al. 2007, p. 37). To some extent, especially concerning equity, this can be connected with individual fairness and group fairness from Dwork et al. (2012) and equal opportunities from Hardt et al. (2016).

Procedural justice means that the process is consistent with all humans, not including bias, accurate, and consistent with the ethical norms (Cropanzano et al. 2007; Leventhal 1980). Consistency plays an essential role in procedural justice, meaning that all employees and all candidates need to receive the same treatment. Additionally, the lack of bias, accuracy, representation of all parties, correction, and ethics play an important role in achieving a high procedural justice (Cropanzano et al. 2007). In contrast, interactional justice is about the treatment of humans, meaning the appropriateness of the treatment from another member of the company, the treatment with dignity, courtesy, and respect, and informational justice (share of relevant information) (Cropanzano et al. 2007).

In general, algorithmic decision-making increases the standardization of procedures, so that decisions should be more objective and less biased, and errors should occur less frequently (Kaibel et al. 2019), since information processing by human raters can be unsystematic, leading to contradictory and insufficient evidence-based decisions (Woods et al. 2020). Consequently, procedural justice and distributive justice are higher using algorithmic decision-making, because the process is more standardized, which still not means that it is without bias.

However, especially in the context of an application or an employee evaluation, it is not only about how fair the procedure itself is (according to fairness measures), but it is also about how people involved in the decision process perceive the fairness of the whole process. Often the personal contact, which characterizes the

interactional fairness, is missing when using algorithmic decision-making. It is difficult to fulfill all three fairness dimensions.

3 Methods

This systematic literature review aims at offering a coherent, transparent, and reliable picture of existing knowledge and providing insights into fruitful research avenues about the discrimination potential and fairness when using algorithmic decision-making in HR recruitment and HR development. This is in line with other systematic literature reviews that organize, evaluate, and synthesize knowledge in a particular field and provide an overall picture of knowledge and suggestions for future research (Petticrew and Roberts 2008; Crossan and Apaydin 2010; Siddaway et al. 2019). To this end, we followed the systematic literature review approach described by Siddaway et al. (2019) and Gough et al. (2017) to ensure a methodical, transparent, and replicable approach.¹

3.1 Search terms and databases

We engaged in an extensive keyword searching, which we derived in an iterative process of search and discussion between the two authors of this study (see “Appendix” for the employed keywords). According to our research question, we first defined individual concepts to create search terms. We considered different terminology, including synonyms, singular/plural forms, different spellings, broader vs. narrow terms, and classification terms of databases to categorize contents (Siddaway et al. 2019) (see Table 2 for a complete list of employed keywords and search strings). Our priority was to achieve the balance between sensitivity and specificity to get broad coverage of the literature and to avoid the unintentional omission of relevant articles (Siddaway et al. 2019).

As the first source of data, we used the social science citation index (SSCI) to ensure broad coverage of scholarly literature. This database covers English-language peer-reviewed journals in business and management. As part of the Web of Knowledge, the database includes all journals with an impact factor, which is a reasonable proxy for the most important publications in the field. We completed our search with the EBSCO Business Source Premier database to add further breadth. Since electronic databases are not fully comprehensive, we additionally searched in the reference section of the considered papers and manually searched for articles (Siddaway et al. 2019).

We considered scholarly articles from a high-quality source of evidence (peer-reviewed and published) journals in English and excluded book reviews, comments, and editorial notes. Moreover, we searched for unpublished articles in conference proceedings from renowned conferences, such as AOM, EURAM, ACM, and IEEE, and contacted the authors to prevent publication bias and to gain further valuable

¹ We thank the anonymous reviewer for this valuable recommendation.

Table 2 Overview of search terms, databases, and results

Search string	Database	Results ^a
TITLE: (“algorithm* OR algorithmic model* OR data-algorithm*OR algorithmic decision-making OR algorithmic decision* OR artificial intelligence OR facial expression tool* OR facial expression processing* OR language processing* OR natural language processing* OR recommender system* OR search engine* OR data*OR data set”)		
TOPIC: (“discrimination* OR discriminat* OR classification* OR “classification problem*” OR “classification scheme*” OR “algorithmic discrimination*” OR “algorithmic bias discrimination*” OR “preventing discrimination*” OR anti-discrimination* OR non-discrimination* OR gender, age, sex, sexism, origin OR “difference* among demographic group*” OR ethic* OR “ethical implication*” OR “data mining discrimination*” OR “unfair treatment*” OR fair* OR unfair* OR “perceived fairness” OR “algorithmic fairness” OR “fairness word*” OR “fairness speech*” OR “fairness recommendation*” OR equal* OR equit* OR inequal* OR “equal opportunit*” OR transparen* OR legal* OR right* OR truth OR impartial* OR correct*OR evaluat* OR judgement* OR “algorithmic judgement*” OR “human judgement*” OR “mechanical judgement*” OR rank* OR rate* OR measure* OR valuation* OR bias* OR “algorithmic bias*” OR “national bias*” OR gender-bias* OR “decision-making bias*” OR “human bias* OR “technical bias*” OR “implicit bias* in algorithm*” OR “dealing with bias*” OR “pattern distortion*” OR pre-justice* OR tendenc* OR prone*OR justice ^b OR adverse impact ^b) AND TOPIC: (“Human Resource*” OR “Human Resource Management” OR Management OR “applicant selection*” OR “employee selection*” OR “algorithm-based HR decision-making” OR “recruitment process* OR “application process*” OR “selection process*” OR recruitment* OR online-recruitment* OR “personnel decision*”, OR “personnel selection*” OR “people analytic*” OR “HR analytic*” OR “job advertisement*” OR “online personalization*”)		
DOCUMENT TYPES = (ARTICLE) AND LANGUAGES = (ENGLISH)	SSCI psychology, psychology experimental, psychology multidisciplinary science, ethics, law, psychology applied, operations research management science, computer science artificial intelligence, computer science interdisciplinary applications, computer science information systems, management, business, behavioral science, social sciences interdisciplinary, sociology, social issues, humanities interdisciplinary	2892 articles
Scholarly (Peer Reviewed) Journals, Academic Journal, Article English	EBSCO Business Source Premier	244 articles

^aResults show the gross hits per search string and database for scholarly articles

^bRobustness check

insights (Siddaway et al. 2019; Lipsey and Wilson 2001; Ferguson and Brannick 2012). In April 2020, this search approach resulted in 3207 articles.

3.2 Screening, eligibility process, and inclusion process

Following this initial identification, we manually screened each article (title and abstract) to evaluate whether its content was fundamental relevant to impact bias, discrimination, or fairness of algorithmic decision-making in HRM, especially in recruitment, selection, development, and training in particular. The process of

relevance screening resulted in 102 articles that were deemed to be substantially relevant.

Second, we conducted the eligibility stage by reading the full text and shifting from sensitivity to specificity. Studies eligible for our review (1) had to be consistent with our definition of algorithmic decision-making as well as with our definitions of fairness, bias, or discrimination (2), and the content had to refer to HRM (3). The list of studies that we excluded at the eligibility stage is available upon request. The two authors independently checked each paper to increase the reliability of the research results. We applied this structured approach to ensure a high level of objectivity.

Afterward, the actual review started, and we synthesized and assessed our findings. We analyzed the material abductively following a set of predefined categories without, however, relying on preexisting codes to extract all relevant information. Analytic categories were, for example, “research design,” “field of the journal,” “research geography,” or “year of publication,” and “key findings.” Again, the authors filled these categories with their inductively generated codes.

Our systematic review used the Preferred Reporting Items for Systematic Reviews (PRISMA) recommendations, including assessment of research content as well as a detailed report of the number of records identified through the search and the number of studies included and excluded in the review. Figure 1 presents a PRISMA flow diagram to provide a succinct summary of the process (Siddaway et al. 2019; Moher et al. 2009).

3.3 Robustness check

We implemented a robustness check to offer a reliable and coherent picture of the discrimination potential and fairness when using algorithmic decision-making in HR recruitment and HR development. With the robustness check, we want to ensure that all relevant articles were included in the literature review. We conducted the robustness check 3 months after the actual search process with two additional keywords, namely: “justice” and “adverse impact” (see Table 2). The search in the database SSCI resulted in 632 articles and the EBSCO search in 690 articles. We manually screened each article (title and abstract) to assess whether the content was essentially relevant to bias, discrimination, or the fairness of algorithmic decision-making in HRM, especially recruitment, selection, training, and development. The majority of articles dealt with the fairness of algorithmic decision-making, but had no reference to HR. After manually screening each article, the process of relevance screening resulted in eight articles for the eligibility stage. We found that no further articles can be included in the literature review by reading the full text. Since out of these eight articles, three articles were already included in the literature review (Lee 2018; Tambe et al. 2019; Yarger et al. 2019), two articles were excluded in the eligibility stage of the initial search process (Hoffmann 2019; Sumser 2017) (no reference to HRM and comment), and the remaining three articles neither discussed fairness nor the HR recruitment and/or HR development context (Varghese et al. 1988; Horton 2017; Gil-Lafuente and Oh 2012). The robustness check verified that the literature review offers a reliable and transparent picture of the current literature

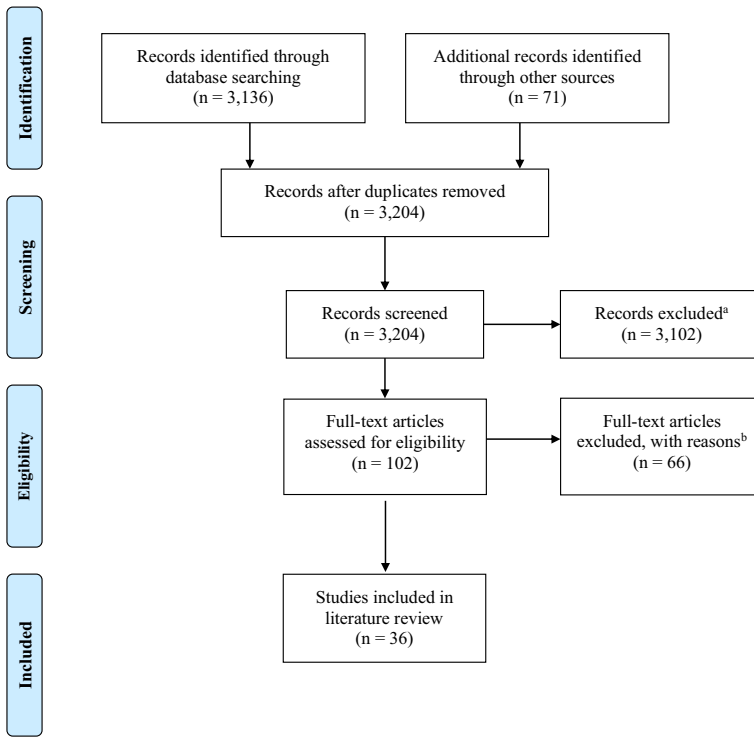


Fig. 1 PRISMA flow diagram illustrating the process. ^aTopic did not fit, mostly no HR and/or fairness, no obvious discrimination context, ^bMostly no HR and/or fairness, no discrimination context after reading the full text or not meeting the inclusion criteria

regarding the discrimination potential and fairness when using algorithmic decision-making in HR recruitment and HR development.

3.4 Limitations of the research process

This approach is not without limitations. First, the reliance on two databases might be regarded as a limitation; however, the approach of selecting two broad and common databases contributed to the validity and replicability of our findings due to the extensive coverage of high-impact, peer-reviewed journals in these databases (Podsakoff et al. 2005). Second, our review focused on two essential HR functions that have severe consequences for individuals and society concerning ethics, namely HR recruitment and HR development. We did not consider other areas of HRM, since the focus of other HR functions is mainly the automation process (e.g., pay or another administrative task). Thus, the situation is different in HR recruitment and HR development, because societal decisions are made, which have crucial consequences for the individual applicants and employees, such as job offer or promotion opportunities. Especially when it comes to decisions about individuals

and their potential, objective and perceived fairness is paramount (Ötting and Maier 2018; Lee 2018).

Moreover, only articles written in the English-language were part of the literature review. Even though this procedure is accepted practice and there is some evidence that including only English articles does not bias the results, it should be noted that non-English articles were not included because English is the dominant language in research (Morrison et al. 2012).

4 Descriptive results

The following section shows the current research landscape. We summarize the main characteristics of the identified articles in Table 3 and present the main findings in Table 4. This table reports the name of authors, year of publication, the main focus of the study (i.e., focus on bias, discrimination, fairness, or perceived fairness), applied method, the field of research, algorithmic decision-making system, HR context (i.e., recruitment- distinguished between recruitment and selection- or development), and the key findings. We analyze the main focus and the key findings of the studies in the following sections. The table is sorted by the focus of the article and whether it is on bias as a trigger for unfairness and discrimination or specifically on fairness and discrimination.

Figure 2 illustrates the distribution of publications over time and the research methods used. The first identified article in our sample of literature was published in 2014. From 2014 to 2016, only a few articles are published per year. From 2017, interest in algorithmic decision-making and discrimination increased notably. As shown in Fig. 2, there was enormous interest in the topic in 2019.

From a methodological perspective, another noteworthy result of this systematic review is the predominance of non-empirical evidence, as Table 3 and Fig. 2 show that the large majority of articles are non-empirical (i.e., conceptual paper, reviews, and case studies). A reason for this is that scientific investigation of discrimination by algorithmic decision-making represents a relatively new topic. However, the number of quantitative papers increased from 2018. Most of the studies focused on bias, discrimination, and objective fairness, while 12 studies examined perceived fairness perceptions of applicants and employees (see Table 1). Furthermore, the majority of studies are located in the area of recruitment and selection, whereby these studies mostly focus on selection. Twelve studies are located in the area of HR development. The majority of studies provided either no geographical specification or were conducted in the USA (see Table 3).

Thirteen articles originate from management, and fourteen articles originate from computer science, four articles originate from law, two from psychology, two from information systems, and one from the behavioral sciences. This distribution illustrates that the field does not have a core in business and management research and is rather interdisciplinary. Nevertheless, the majority of articles originating from management were published in high-ranked journals, such as *Journal of Business Ethics*, *Human Resource Management Review*, *Management Science*, *Academy of Management Annals*, and *Journal of Management*. The majority of these studies

Table 3 Overview of studies

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Naim et al. (2016)	The automated analysis of facial expressions, language, and prosodic information of interviewees in a job interview	B	Empirical-quantitative; analysis of interviews	Computer science	NLP, FEP	Selection	<p>Recommends to speak more fluently, use less filler words, and smile more</p> <p>Shows that the students who were rated highly while answering the first interview question were also rated highly overall (i.e., first impression matters)</p>	USA
Cheng and Hackett (2019)	A critical review of algorithms in HRM	B	Empirical-qualitative; single case study; review	Management	General	Recruitment and selection	<p>Organizations have to increase the perceived authenticity of algorithms</p> <p>Need to evaluate algorithms from a research perspective</p> <p>Lack between practice and research</p>	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Mann and O'Neil (2016)	Explains why algorithms are not neutral and offers some implications to reduce the risk of biases of algorithmic decision-making in the hiring process	B, D	Non-empirical; conceptual paper	Management	General, hiring algorithms	Selection	Algorithms reflect human biases and prejudices that lead to machine learning mistakes and misinterpretations Bias and prejudice cannot be completely eliminated from hiring HR professionals must consider the consequences of these systems and ensure they always reflect the best human intentions	USA
Kim (2017)	Examines the use of classification schemes/data algorithms in terms of personnel decisions Shows limitations in existing law and improvement proposals	B, D	Non-empirical; conceptual paper	Law	General, classification	Recruitment and selection, training and development	Because of the nature of data mining techniques, employer reliance on these tools poses novel challenges to workplace equality	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Rosenblat et al. (2016)	Examines discrimination thorough evaluation in the case of Uber drivers	B, D	Empirical-qualitative; case study	Computer science	Evaluation systems	Development	The need to exercise quality control over a large disaggregated workforce may permit the continued use of rating Systems under existing employment discrimination law	Not specified
Savage and Bales (2017)	Examines how video game algorithms are incorporated into the job hiring process Shows the debate over whether these algorithms discriminate	B, D	Non-empirical; conceptual paper	Law	Gamification	Selection, development	Video games in initial hiring stages can permit non-discriminatory evaluation of all the candidates	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Williams et al. (2018)	Examines discrimination through the use of algorithms in decision-making processes Proposes strategies for the prevention of discrimination	B, D	Empirical-qualitative; multiple case studies	Information	General	Recruitment and selection	Algorithmic prediction can include injustices The predictions have to be checked for bias and should be corrected to avoid discrimination	USA
Lambrecht and Tucker (2019)	Examines gender bias in delivery of job ads Conducts field test of how an algorithm delivered ads promoting job opportunities in the science, technology, engineering and math fields	B, D	Empirical-quantitative; field test	Management	Recommender systems	Recruitment	Fewer women saw STEM ads than men An algorithm that simply optimizes cost-effectiveness in ad delivery will deliver ads that were intended to be gender neutral in an apparently discriminatory way	191 countries

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Sajjadiami et al. (2019)	Prediction of future work outcomes such as voluntary turnover, involuntary turnover, and value-added based on work history criteria (work experience relevance, tenure history, attributions for previous turnover)	B, F	Empirical-quantitative, statistical tests; machine learning	Psychology	General	Selection	Algorithmic methods are often calibrated exclusively to a specific application pool Pairing established theory leads to better results than unique word applications	USA
Yarger et al. (2019)	Critical analysis of talent acquisition software and its potential for fostering equity in the hiring process for underrepresented IT professionals	B, F	Non-empirical; conceptual paper	Information	General	Recruitment and selection	Human expertise is still necessary Even well-intentioned algorithms are not neutral and should be audited for morally and legally unacceptable decisions	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Raghavan et al. (2020)	Documents and analyzes the claims and practices of companies offering algorithms for employment assessment	B, F	Empirical-qualitative; case study	Computer science	Employment assessment	Selection	Target variables and training data: most of the vendors offer customizable assessment Validation: vendor's websites often do not clarify whether they validate their models	Several countries
Sánchez-Monedero et al. (2020)	Examines how three automated hiring systems can be understand and attempt to mitigate bias and discrimination in the UK	B, F	Empirical-qualitative; case study	Computer science	Employment assessment	Selection	Often lack of information on how the system worked Claims and validation are often vague	UK

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Stone et al. (2015)	Reviews several of the primary forces that are presenting challenges for HR research and practice Tries to answer the question whether eHRM influences organizational effectiveness and whether it enables organizations to achieve their HR goals	B, PF	Non-empirical; conceptual paper	Management	Several electronic HRM tools	Recruitment and selection, development	There are still a number of questions about whether these new systems enable organizations to achieve their primary HR goals	USA
Woodruff et al. (2018)	Explores how members of potentially affected communities in the USA feel about algorithmic fairness	B, PF	Empirical; qualitative; workshop; interviews	Computer science	General	Recruitment and selection	Concept of algorithmic fairness is largely unfamiliar Learning about algorithmic (un)fairness elicited negative feelings Company handling of algorithmic fairness interacts significantly with user trust	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Leicht-Deobald et al. (2019)	Identifies challenges arising from algorithm-based HR decision-making Analyzes how algorithm-based HR Decision-making may influence employees' personal integrity, and actions	B, PF	Non-empirical; conceptual paper	Management	General	Development	Algorithm-based decision-making can help monitor employees more effectively but can be ethically problematic Suggests four mechanisms to reduce negative consequences	Not specified
Tambe et al. (2019)	Identifies challenges in using data science techniques for HR tasks Proposes practical responses to these challenges based on the principles of causal reasoning, randomization and experiments, and employee contribution	B, PF	Empirical; qualitative; workshop and survey	Management	General	Recruitment and selection	Four challenges identified: complexity of HR phenomena, constraints imposed by small data sets, accountability questions associated with fairness and other ethical and legal constraints, and possible adverse employee reactions	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Rosenblat et al. (2014)	Focuses on the new tool's employers use to sift through job applications Addresses issues of privacy, fairness, transparency, accuracy, and inequality under the rubric of discrimination	B, D, F	Non-empirical; literature review	Computer science	General	Recruitment and selection	Employers potentially have access to more comprehensive electronic profiles on job candidates than has been traditionally available to them, which can expose job candidates to a greater scrutiny of their personal lives	Not specified
Burdon and Harpur (2015)	Examines potential for discriminatory practices to develop through information infrastructures in which unfairness and discrimination are embedded into the prescriptive processes and infrastructures of talent analytics	B, D, F	Non-empirical; conceptual paper	Law	General	Selection, development	Processes of predictive segmentation can produce inequalities through the segmentation of employee groupings based on unintuitive attributes that are ever-changing	Australia

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Persson (2016)	Analyzes the problems of implicit bias in algorithms regarding recruitment processes, discrimination, and unfairness by using algorithms Shows possible solutions	B, D, F	Empirical-qualitative; explorative analysis	Computer science	Data mining, profiling	Recruitment and selection	Employers might miss the best candidates, as the employed algorithms are tuned with limited and outdated data The risk of directly or indirectly discriminating candidates exists	Not specified
Vasconcelos et al. (2017)	Proposes a structured approach to mitigate discrimination and unfairness caused by bias in AI systems in hiring decision scenarios	B, D, F	Non-empirical; conceptual paper	Computer science	General, hiring algorithms	Selection	Points out connections between bias of AI and the problem of induction Shows that there is a logical theory of preferences	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Chen et al. (2018)	Investigates gender-based inequalities in the context of resume search engines, discrimination	B, D, F	Empirical-quantitative; statistical tests	Computer science	Search engines	Recruitment	Individual fairness: even when controlling for all other visible candidate features, there is a slight penalty against female candidates Group fairness: 8.5–13.2% of job title/city pairs show statistically significant group unfairness	USA
Bogen (2019)	Analysis of predictive tools across the hiring process to clarify what hiring algorithms do, and where and how bias can enter into the process	B, D, F	Non-empirical; conceptual paper	Management	General, hiring algorithms	Recruitment and selection	Most hiring algorithms will drift toward bias by default Potential to help reduce interpersonal bias should not be discounted Only tools that proactively tackle deeper disparities will offer any hope that predictive technology can help promote equity, rather than erode it	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Simbeck (2019)	Discusses the ethical implications of the application of sophisticated analytical methods to questions in HR management	B, D, F	Non-empirical; conceptual paper	Management	General	Recruitment and selection, development	The rising data on employees will lead to an unprecedented transparency of employees Proposes to transfer key ethical concepts from medical research, artificial intelligence, learning analytics, and coaching to HR analytics	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Kellogg et al. (2020)	Analyzes how the implementation of algorithmic technologies in organizations may reshape organizational control	B, D, F	Non-empirical; literature review	Management	Algorithmic recommending, restricting, recording, rating, replacing, and rewarding	Development	Algorithmic control in the workplace operates through six main mechanisms: restricting, recommending, recording, rating, replacing, rewarding The technical capabilities of algorithmic systems facilitate a form of rational control that is distinct from control used by employers for the past century Employees are individually and collectively resisting algorithmic control through a set of emerging tactics	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Chamorro-Premuzic et al. (2017)	Reviews three innovations that have the potential to revolutionize the way organizations identify, develop, and engage talent and are emerging as tools used by practitioners and firms	D	Non-empirical; conceptual paper	Behavioral science	Machine-learning algorithms, social sensing technology, gamified Assessment tools	Recruitment and selection, development	Talent analytics represent a more data-driven and evidence-based approach than human intuition	Not specified
Kim and Scott (2018)	Examines discrimination in the online application process Develops mechanisms to check when discrimination occurs in online recruitment	D	Empirical-qualitative	Law	Recommender systems	Recruitment	It is not certain whether current law is adequate to reach all forms of targeted Recruitment with significant discriminatory effects	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Cappelli (2019)	Presents problems which occur because of the rise of data science in hiring processes by using driven algorithms to find and assess job candidates	D	Non-empirical; illustrative case	Management	General	Selection	Possible solution might be tracking the percentage of openings filled from within, requiring that all openings be posted internally and recognizing the costs of outside hiring Ways to revamp the hiring process could be among others not to post “phantom jobs“, designing jobs with realistic requirements	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Lee et al. (2015)	Explores the impact of algorithmic, data-driven management in the context of new ride sharing services Highlights opportunities and challenges in designing human-centered algorithmic work assignment, information, and evaluation	F, PF	Empirical-qualitative; semi-structured interviews; post analysis	Computer science	Evaluation process	Development	Highlights how transparency of algorithmic assignment influences employee's cooperation, work strategy, and workload creation These numeric systems that made drivers accountable for all interactions were sometimes seen as unfair and ineffective and created negative psychological feelings in drivers	USA

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
McCarthy et al. (2017)	Comprehensive review of research on applicant reactions to selection procedures published since 2000 (including algorithmic based selection tools)	PF	Non-empirical; narrative review	Management	Several selection tools	Selection	The field of applicant reactions has advanced considerably and made substantial and meaningful contributions to theory, methods, and practice of recruitment and selection over the last 15 years	Not specified
Langer et al. (2018)	Analyzes the role of computer experience and information on applicant reaction towards novel technologies for personnel selection	PF	Empirical-quantitative; quasi-experimental design	Computer science	Several selection tools	Selection	Computer experience did not affect the candidate reaction	Not specified

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Lee (2018)	Explores the perceptions of algorithmic management by conducting an online experiment using four managerial decisions The decision-maker (algorithmic or human) was manipulated, and measured perceived fairness, trust, and emotional response	PF	Empirical-quantitative; experimental design	Computer science	General	Selection, development	People perceive algorithmic decisions as less trustworthy and more likely to evoke negative emotion for tasks that people think require uniquely human skills	USA
Ötting and Maier (2018)	Examines the effects of procedural justice and the type of decision agent on employee behavior and attitudes	PF	Empirical-quantitative; experimental design	Computer science	General	Development	Significant effects of procedural justice on employee behavior and attitudes, confirming the importance of procedural justice at the workplace for both human and system decision agents	Germany

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Kaibel et al. (2019)	Analysis of perception of algorithm-based decisions versus human-based decisions in recruitment Role of interindividual differences	PF	Empirical-quantitative; Experimental design	Management	Automated screening algorithm, digital interview	Selection	Applicants perceived the algorithmic selection as more consistent, less personable The organizational attractiveness is lower with algorithm-based decision	Germany, USA
Langer et al. (2019)	Comparison of highly automated interviews to a video conference interview	PF	Empirical-quantitative; experimental design	Psychology	Asynchronous video interviews	Selection	Highly automated interviews are less accepted than video conference interviews from the perspective of applicant and need to be more transparent	Germany
Suen et al. (2019)	Investigates the social impacts of using synchrony and AI decision agents in video interviews	PF	Empirical-quantitative; experimental design	Computer science	Asynchronous video interviews	Selection	Job applicants had indistinguishable fairness perceptions regarding synchrony and decision agent (human vs. AI)	China

Table 3 continued

Author(s), year	Main focus	B/D/F/PF	Method	Field of research	System	Recruitment, selection, development, training	Key findings	Geography
Leclercq-Vandermoortte (2017)	Investigates the moral and ethical implications of emerging forms of control that have developed along with the use of modern ubiquitous information technology in the workplace	Not specified	Non-empirical; case study	Management	General	Development	Ethical issues are the ambivalence of the use of ubiquitous IT at work, subtlety of the control exerted by ubiquitous IT, invasiveness of ubiquitous IT, and self-reinforcement of ubiquitous IT-based control	France

B bias, *D* discrimination, *F* fairness, *PF* perceived fairness

Table 4 Types of AI application, bias, research gaps, and research implications

	HR recruitment		HR development
	Recruitment	Selection	
HR tools	Search engines Recommender systems	Analysis of CV, résumé, and references Gamification Algorithmic video- and telephone analysis	Data profiling Evaluation Interviews Recommender systems
AI tools	Collaborative filtering	Text mining, NLP, FEP	NLP, FEP, data mining, collaborative filtering, classification
Bias	Yes	Yes	Yes
Perceived fairness	No	Yes	Yes
Research implications	Need for empirical-quantitative studies Research about the perceived fairness	Research on AI in several steps of the selection process Research on AI as decision support	Need for empirical-quantitative studies Research on the perceived fairness
Exemplary research questions	How fair do applicants perceive search engines and recommender systems?	How can established selection techniques be transformed into algorithmic measurements? To what extent can algorithmic de-bias strategies be applied? Which type of algorithmic decision tool do applicants prefer? Does more information about the algorithmic decision-making process influences the fairness perception of the applicants? How do candidates react when AI is interfered in several selection process steps? What is the reaction if the algorithm only supports the process, but the final decision remains human? What is the difference in reliability and validity between AI decision-makers and human raters?	To what extent can algorithmic de-bias strategies be applied? What are ways to avoid bias? How fair do employees consider algorithmic decision-making? What measures can be taken to ensure that employees perceive algorithmic decision-making as fair?

CV curriculum vitae, *FEP* facial expression processing, *NLP* natural language processing, *ML* machine learning

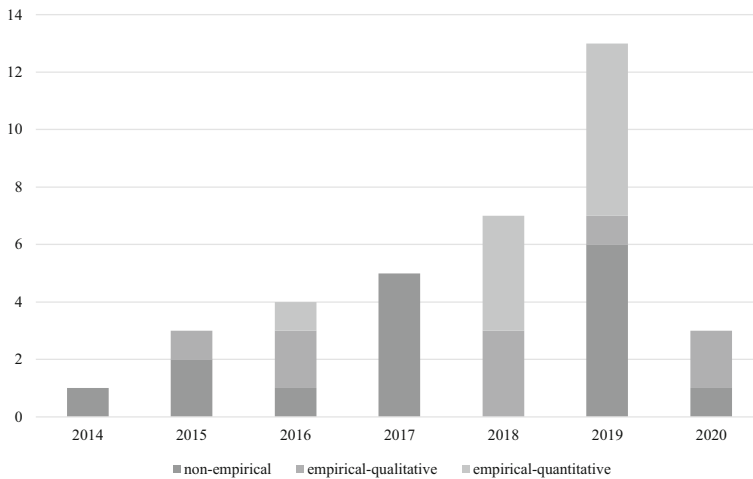


Fig. 2 Distribution of publications over time and research methods. Data on 2020 research articles are based on our database search until April 2020

were published in 2019, which stresses the importance of fairness and discrimination as a recent topic in the management and HRM literature.

Our results suggest there is still room for academic researchers to complement the literature and discussion on algorithmic decision-making and fairness. In the following, we introduce some algorithmic decision tools used in HR recruitment and HR development and their potential for discrimination.

5 Types of algorithmic decisions and applications in HR

5.1 HR recruitment

In the following, we present some examples of algorithmic decision-making applications in HR recruitment and their fairness. We distinguish between recruitment (i.e., finding a candidate) and selection (i.e., selecting among these candidates), which is considered as part of the recruitment process, because, in these two different stages, companies use different algorithmic decision tools.

Firms increasingly rely on social media platforms and digital services, such as Facebook, Instagram, LinkedIn, Xing, Monster, and CareerBuilder, to advertise job vacancies and to find well-fitting candidates (Burke et al. 2018; Chen et al. 2018). These digital services are called recommender systems and search engines and use algorithmic decision-making tools to recommend suitable candidates to recruiters and suitable employers to candidates (Chen et al. 2018). To propose individual recommendations, recommender systems take advantage of different information sources. Based on users' descriptions, prior choices, and the behavior of other similar users, the recommender system proposes ads aiming to match recommendations and user preferences (Burke et al. 2018; Simbeck 2019). However, it is a

multifaceted concept, not only the users (here: job seekers) need to be considered, but also stakeholders (Burke et al. 2018). Hiring platforms, such as Xing and LinkedIn, already implement predictive analytics. Their algorithms go through thousands of job profiles to find the most eligible candidate for a specific job and recommend this candidate to the recruiter (Carey and Smith 2016). Firms also examine data about job seekers, analyze them based on past hiring decisions, and then recommend only the applications that are a potential match (Kim 2016). Consequently, firms can more precisely target potential candidates. These predictions based on past decisions can unintentionally lead to companies using job advertisements that strengthen gender and racial stereotypes, because if, for example, in the past, more males were selected for high position jobs, the advertisement is consequently shown to more males (historical bias). Thus, tension exists between the goals of fairness and those of personalization (Burke et al. 2018).

In a non-empirical paper analyzing predictive tools in USA, Bogen (2019) gives a prime example of algorithmic discrimination against other genders by demonstrating that algorithms extrapolate based on patterns of the provided data. Thus, if recruiters contacted males more frequently than females, the recommendation will be to show job ads more often to males. An explanation could be that males are more likely to click on high-paying job ads, and consequently, the algorithm learns from this behavior (Burke et al. 2018).

Another example showed that targeted ads on Facebook were predominately shown to females (85%), while jobs advertised by taxi companies were shown mainly to males (Bogen 2019). In their field test of how an algorithm delivered ads promoting job opportunities in the STEM fields, Lambrecht and Tucker (2019) found in an empirical-quantitative field test among 191 countries that online job advertisements in the science, technology, engineering, and math sector were more likely shown to males than females. This gender bias in the delivery of job ads occurs, because even if the job advertisement should be delivered explicitly gender neutral, an algorithm that optimizes cost-effectiveness in ad delivery would deliver ads discriminatorily due to crowding out (Lambrecht and Tucker 2019).

Platforms, such as Google, LinkedIn, and Facebook, offer advertisers the possibility to target viewers based on sensitive attributes to exclude some job seekers depending on their attributes (Kim and Scott 2018). For instance, Facebook let firms choose among over 100 well-defined attributes (Ali et al. 2019). In this case, humans interact and determine the output strategically (intentional discrimination). For example, through their selection of personal traits, older potential candidates are excluded from seeing the job advertisement. Companies make use of targeted ads to attract job seekers who are most likely to have relevant skills, while recommender systems can reject a large proportion of applicants (Kim and Scott 2018). Even if companies chose their viewers by relying on attributes that appear to be neutral, these attributes can be closely related to protected traits, such as ethnicity, and could allow biased targeting. Often, bias in recommender systems can occur unintentionally and rely on attributes that are not obvious (Kim and Scott 2018). Kim and Scott (2018) analyzed in an empirical-qualitative paper that due to spillover effects, it is more costly to serve ads to young females, because women on Facebook are known to be more likely to click on ads (Kim and Scott 2018). Hence,

algorithms that optimize cost efficiency may deliver ads more often to males, because they are less expensive than females (Kim and Scott 2018). In summary, these three studies based on non-empirical, empirical-qualitative, and empirical-quantitative evidence show that historical biases and biases caused by cost-effectiveness reasons occur in HR recruitment and selection.

With the help of search engines, recruiters proactively search for candidates who use employment services on keywords and filters (Chen et al. 2018). The algorithm rates applicants; consequently, the recruiter sees and more likely clicks on those at the top. These rankings often take demographic features (e.g., name, age, country, and education level) into account, and this can yield a disadvantage for some candidates (Bozdog 2013; Chen et al. 2018). Other features are, for example, the locations, previous search keywords, and the recent contacts in a user's social network. These service sites do now allow recruiters to filter search results by demographics (e.g., gender, age, and ethnicity). Nonetheless, these variables exist indirectly in other variables, such as years of experience as an indicator of age (Chen et al. 2018). With the help of statistical tests and data on 855,000 USA job candidates (search results for 35 job titles across 20 USA cities), Chen et al. (2018) revealed in an empirical-qualitative single case study and review that the search engines provided by Indeed, Monster, and CareerBuilder discriminate against female candidates to a lesser extent.

5.2 HR selection

Striving for more efficiency due to time and cost pressures and limited resources by simultaneously managing a large number of applications are among the main reasons for the increasing use of algorithmic decision-making in the selection context (Leicht-Deobald et al. 2019). Organizations are increasingly using algorithmic decision tools, such as CV and résumé screening, telephone, or video interviews, providing an algorithmic evaluation (Lee and Baykal 2017; Mann and O'Neil 2016) before conducting face-to-face interviews (Chamorro-Premuzic et al. 2016; van Esch et al. 2019).

One possibility for using algorithmic decision-making in selection is the analysis of the CV and résumé, with candidates entering their CVs or job preferences online, and this information is subject to algorithmic analysis (Savage and Bales 2017). Yarger et al. (2019) conceptually analyzed the fairness of talent acquisition software in the USA and its potential to promote fairness in the selection process for underrepresented IT professionals. The authors argue that it is necessary to audit algorithms, because they are not neutral. One prominent example is the CV screening tool of Amazon, which was trained on biased historical data that led to a preference for male candidates based on the fact that, in the past, Amazon hired more often males as software engineers as females and the algorithm has been trained based on these data (historical bias) (Dastin 2018). Yarger et al. (2019) suggest removing sources of human bias such as gender, race, ethnicity, religion, sexual orientation, age, and information that can indicate membership in a protected class. Text mining is often the foundation for the screening of CVs and résumés, an approach to characterize and transform text using the words themselves as the unit

of analysis (e.g., the presence or absence of a specific word of interest) (Dreisbach et al. 2019).

Besides words, also certain criteria, such as gender and age, play an important role when the training of the algorithm is based on data which has exhibited a preference for males, females, or younger people in the past. Thus, the algorithm eliminates highly qualified candidates who do not present selected keywords or phrases or who are of a specific age or gender (Savage and Bales 2017). Applying machine learning and statistical test in an empirical-quantitative setting, Sajjadi et al. (2019) suggest analyzing and developing interpretable measures that are integrated with a substantial body of knowledge already present in the field of selection and established selection techniques rather than relying on the unique word application. One example is to pair job titles with job analysts' rankings of task requirements in O*NET to have more valid predictions.

Qualifications that cannot be observed through analyzing the résumé can be analyzed by means of gamification. Here, applicants take quizzes or play games, which allow an assessment of their qualities, work ethic, problem-solving skills, and motivation. Savage and Bales (2017) argue in a non-empirical conceptual paper that video games in initial hiring stages permit a non-discriminatory evaluation of all candidates, because it eliminates the human bias, and only the performance in the game counts.

Another application of algorithmic evaluation and widely used by companies is video and telephone analyses (Lee and Baykal 2017). Candidates answer several questions via video (HireVue OnDemand 2019) or telephone (Precire 2020; 8andAbove 2020), and their responses are analyzed algorithmically (Guchait et al. 2014). With the help of sensor devices, such as cameras and microphones, human verbal and nonverbal behavior is captured and analyzed by an algorithm (Langer et al. 2019). AI tools for identifying and managing these spoken texts and facial expressions are natural language processing (NLP) and facial expression processing (FEP). “[...] NLP is a collection of syntactic and/or semantic rule- or statistical-based processing algorithms that can be used to parse, segment, extract, or analyze text data” (Dreisbach et al. 2019, p. 2). Word counts, topic modeling, and prosodic information, such as pitch intention and pauses, will be extracted by an algorithm, resulting in the applicant's personality profile, e.g., Big Five. FEP analyzes facial expressions, such as smiles, head gestures, and facial tracking points (Naim et al. 2016).

During the asynchronous video interview, applicants record their answers to specific questions and upload them to a platform. In the case of telephone interviews, the applicant speaks with a virtual agent (Precire 2020). Companies make use of ML algorithms to predict which candidate is best suited for a specific job. For example, HireVue provides a video-based assessment method that uses NLP and FEP to assess candidates' stress tolerance, their ability to work in teams, or their willingness to learn. As a result of technological advances, it is now possible to create a complete personal profile. Based on a case study, Raghavan et al. (2020) analyzed the claims and practices of companies offering algorithms for employment assessment and found that the vendors, in general, do not particularly reveal much about their practices; thus, there is a lack of transparency in this area.

Turning the perspective from the employer to the candidates, especially the perceived fairness of the candidates, plays an essential role in recruitment outcomes (Gilliland 1993). Using a between-subject online experiment, Lee (2018) discovered that people perceive human decisions to be fairer than algorithmic decisions in hiring tasks. People think that the algorithm lacks the ability to discern suitable applicants, because the algorithm makes judgments based on keywords and does not take qualities that are hard to quantify into account. Participants do not trust the algorithm, because it lacks human judgment and human intuition. Contrasting findings are found in Suen et al.'s (2019) empirical-quantitative study comparing synchronous videos to asynchronous videos analyzed by means of an AI; they conclude that the videos analyzed by means of an AI did not negatively influence perceived fairness in their Chinese sample.

Unlike the other studies, in an online experiment, Kaibel et al. (2019) recently analyzed the perceived fairness of two different algorithmic decision tools, namely initial screening and digital interviews. Results show that algorithmic decision-making negatively affects personableness and the opportunity to perform during the selection process, but it does not affect the perceived consistency. These relationships are moderated by personal uniqueness and experienced discrimination.

5.3 HR development

Research on fairness of algorithmic decision-making and HR development is still in its infancy, since most existing studies focus on the fairness of the recruitment process.

Companies increasingly rely on algorithmic decision-making to quantify and monitor their employees (Leicht-Deobald et al. 2019). Personal records and internal performance evaluation are documented in firm systems. Identifying knowledge and skills is a major aim of algorithmic decision-making in HR development (Simbeck 2019). Other goals are workforce forecasts (retention, leaves) and comprehension of employee's satisfaction indicators (Simbeck 2019; Silverman and Waller 2015). Typical data stored in HR information systems include information about the employees hired, the employee's pay and benefits, hours worked, and sometimes various performance-related measures (Leicht-Deobald et al. 2019). Personal data, such as the number and age of children, marital status, and health information, are often available for the HR function (Simbeck 2019). Companies that offer employee engagement analytics, performance measurement, and benchmarking include, for example, IBM (Watson Talent Insights), SAP (Success Factors People Analytics), and Microsoft (Office 365 Workplace Analytics). These algorithmic decision tools offer opportunities to organize the employee's performance more effectively, but they also associated with certain risks. Since HR development is about assessing and improving the performance of the employees by applying algorithmic decision-making, there are several overlaps with HR recruitment. While HR recruitment focuses on predicting the performance of candidates, HR development focuses on developing existing employees and talents. Nevertheless, the tools used are quite similar.

One of the methods that is used is data profiling, which is a special use of data management. It aims to discover the meaningful features of data sets. The company is provided with a broad picture of the data structure, content, and relationships (Persson 2016). One company, for example, observed that the distance between the workplace and home is a strong predictor of job tenure. If a hiring algorithm relied on this aspect, discrimination based on residence occurs (Kim 2016). Additionally, NLP is also used in the HR development. To identify skills and to support career paths, some companies conduct interviews with their employees to create a psychological profile (e.g., personality or cognitive ability) (Chamorro-Premuzic et al. 2016).

Another approach is evaluation. For example, Rosenblat and Stark (2016) examined in a case study the evaluation platform of the American passenger transport mediation service company Uber and found that discrimination exists in the evaluation of drivers. Uber tracks employees' GPS positions and has acceleration sensors integrated into the driver's version of the Uber app to detect heavy braking and speeding (Prassl 2018). Females are paid less than males, because they drive slower. Consequently, the algorithm calculates a lower salary due to slower driving for the same route.

To evaluate and promote employees, organizations increasingly rely on recommender systems. For example, IBM offers IBM Watson Career Coach, which is a career management solution that advises employees about online and offline training based on their current job and previous jobs within the company and based on the experiences of similar employees (IBM 2020). The pitfalls with respect to recommender systems, as mentioned earlier, also apply in the development.

Regarding the perceived fairness, in an empirical-quantitative online experiment Lee (2018) analyzed the fairness perception of managerial decisions (using a customer service call center that uses NLP to evaluate the performance), whereby the decision-maker was manipulated. Performance evaluations carried out by an algorithm are less likely to be perceived as fair and trustworthy, and at the same time, they evoke more negative feelings than human decisions.

6 Discussion

This paper aimed at raising awareness of the potential problems regarding discrimination, bias, and unfairness of algorithmic decision-making in two important HR functions dealing with an assessment of individuals, their potential, and their fit to the organization. While previous research highlighted the organizational advantages of algorithmic decision-making, including cost savings and increased efficiency, the possible downsides in terms of biases, discrimination, and perceived unfairness have found little attention in HRM, although these issues are well known in other research areas. By linking these research areas with HR recruitment and HR development and identifying important research gaps, we offer fruitful directions for future research by highlighting areas where more empirical evidence is needed. Consequently, a major finding that emerges from our literature

review is the need for more quantitative research on the potential pitfalls of algorithmic decision-making in the field of HRM.

Companies implement algorithmic decision-making to avoid or even overcome human biases. However, our systematic literature review shows that algorithmic decision-making is not a panacea for eliminating biases. Algorithms are vulnerable to biases in terms of gender, ethnicity, sexual orientation, or other characteristics if the algorithm builds upon inaccurate, biased, or unrepresentative input and training data (Kim 2016). Algorithms replicate biases if the input data are already biased. Consequently, there is a need for transparency; employees and candidates should have the possibility to understand what happens within the process (Lepri et al. 2018).

Moreover, organizations need to consider the perceived fairness of employees and applicants when using algorithmic decision-making in HR recruitment and HR development. For companies, it is difficult to satisfy both computational fairness from the computer science, which is defined by rules and formulas, and perceived fairness from the management literature that is subjectively felt by potential and current employees. To fulfill procedural justice and distributive justice, it is important for organizations to reduce or avoid all types of biases and to achieve subjective fairness, such as individual fairness, group fairness (Dwork et al. 2012), and equal opportunity (Hardt et al. 2016). Companies need to continuously enhance the perceived fairness of their HR recruitment and selection and HR training and development process to avoid adverse impacts on the organization, such as diminishing employer attractiveness, employer image, task performance, motivation, and satisfaction with the processes (Cropanzano et al. 2007; Cohen-Charash and Spector 2001; Gilliland 1993).

With regard to fairness perceptions, it appears to be beneficial that humans make the final decision if the decision is about the potential of employees or career development (Lee 2018). At first glance, this partially contradicts previous findings that the automated evaluation seems to be more valid, since human raters may evaluate candidates inconsistently or without proper evidence (Kuncel et al. 2013; Woods et al. 2020). However, while people accept that an algorithmic system performs mechanical tasks (e.g., work scheduling), human tasks (e.g., hiring, work evaluation) should be performed by humans (Lee 2018). Reasons for the lower acceptance of algorithms in judging people and their potential are multifaceted. The usage of this new technology in HRM, combined with a lack of knowledge and transparency about how the algorithms work, increases emotional creepiness (e.g., Langer et al. 2019; Langer and König 2018) and decreases interpersonal treatment and social interactions (e.g., Lee 2018) as well as fairness perceptions and the opportunity to perform (e.g., Kaibel et al. 2019). To overcome these adverse impacts of algorithmic decision-making in HRM, companies need to promote their usage of algorithms (van Esch et al. 2019) and make the processes more transparent of how algorithms are supporting the decisions of humans (Tambe et al. 2019). This might help to create HR systems in recruitment and career development that are both valid and perceived as fair. Nevertheless, a fruitful research avenue could be to examine how companies should communicate or promote their usage of algorithms

and whether employees and applicants accept a certain degree of algorithmic aid in human decision-making.

In summary, companies should not solely rely on the information provided by algorithms or even implement automatic decision-making without any control or auditing by humans. While some biases might be more apparent, implicit discrimination of less apparent personal characteristics might be more problematic, because such implicit biases are more difficult to detect. In the following, we outline theoretical and practical implications as well as future research directions.

6.1 Theoretical implications and future research directions

This review reveals that current knowledge on the possible pitfalls of algorithmic decision-making in HRM is still in an early stage, although we recently identified increased attention to fairness and discrimination. Thus, the question arises about what the most important future research priorities are (see Table 4 for exemplary research questions). The majority of studies which we found concerning fairness and discrimination were non-empirical. One reason for the paucity of empirical research could be that algorithmic decision-making is a recent phenomenon in the field of HR recruitment and HR development, which has not yet received much attention from management scholars. Consequently, there is a need for more sophisticated, theoretically, quantitative studies, especially in HR recruitment and HR development, but also in HR selection. In this regard, a closer look reveals that the majority of current research focuses on HR selection. However, also for HR selection, only one or two studies per tool addressed fairness or perceived fairness. In contrast, fairness perceptions and biases in HR recruitment and HR development receive little attention (see Table 3).

The discussion on what leads to discrimination and its avoidance seems to be a fruitful research avenue. Notably, the different types of algorithmic bias (see Sect. 2.2) that can lead to (implicit) discrimination and unfairness need to be considered separately. The existing studies mainly discuss bias, unfairness, and discrimination in general, but rarely delve into detail by studying what kind of bias occurred (e.g., historical bias or technical bias). Similarly, several studies distinguished between mathematical fairness and perceived fairness, but did not take a closer look at individual fairness, group fairness, or equal opportunity (see Sect. 2.3).

Another prospective research area focuses on the difference in reliability and validity between AI decision-makers and human raters (Suen et al. 2019). Many studies found that an algorithm could be discriminatory, but the question remains whether algorithms are fairer than humans are. However, this is important to address to achieve the fairest possible decision-making process.

Another research avenue for new tools in HR recruitment and HR development focuses on the individuals' perspective and acceptance of algorithmic decision-making. Only a few studies have examined the subjective fairness perceptions of algorithmic decision-making in the HRM context. Thus, the way employees and applicants perceive decisions made by an algorithm instead of humans is not fully exploited (Lee 2018). In HR selection, a few studies have analyzed the perceived

fairness. However, our systematic review underlines the recent calls by Hiemstra et al. (2019) and Langer et al. (2018) for additional research to fully understand the emotions and reactions of candidates and talented employees when using algorithmic decision-making in HR recruitment or HR development processes. Emotions and reactions can have important negative consequences for organizations, such as withdrawal from the application process or job turnover (Anderson 2003; Ryan and Ployhart 2000). In general, knowledge about applicants' reactions when using algorithmic decision-making is still limited (van Esch et al. 2019). Previous studies analyzed a single algorithmic decision tool [see Kaibel et al. (2019) for a recent exception]. Consequently, there is a need to examine applicants' acceptance of algorithmic decision-making within the steps of the recruitment and selection process (e.g., media content and recruitment tools on the employer's webpage, recommender systems in social media, screening and preselection, telephone interview, and video interview).

Although there is some evidence that candidates react negatively to a decision made by an algorithm (i.e., Kaibel et al. 2019; Ötting and Maier 2018; Lee 2018), more research is needed on individuals' acceptance of algorithms if algorithms support the decisions by humans. Moreover, additional insights are needed into whether transparency and more information about the algorithmic decision-making process positively influences the fairness perception (Hiemstra et al. 2019). Finally, while we found many studies examining the fairness perception of applicants (i.e., potential employees), the perspective of current employees on algorithmic decision-making is still neglected in HRM research. Besides the threat of job loss due to digitalization and automation, the question of how algorithms might help to assess, promote, and retain qualified and talented employees remains important and will become more important in the next decade. Thus, fairness and biases perceived by current employees offer yet another fruitful research avenue in HR development.

6.2 Practical implications

Given that in many companies, the HR function has the main responsibility for current and potential employees, our literature review shows that HR managers need to be careful about implementing algorithmic decision-making, respecting privacy and fairness concerns, and monitoring and auditing the algorithms that are used (Simbeck 2019). This is accompanied by an obligation to inform employees and applicants about the usage of the data and the potential consequences, for example, forecasting career opportunities. Since the implementation of algorithmic decision-making in HRM is a social process, employees should actively participate in this process (Leicht-Deobald et al. 2019; Friedman et al. 2013; Tambe et al. 2019). Moreover, applicants and employees must have the opportunity to not agree with the proceedings (Simbeck 2019). A first step would be to implement company guidelines for the execution and auditing of algorithmic decision-making and transparent communication about data usage (Simbeck 2019; Cheng and Hackett 2019).

If companies implement an algorithm, the responsibility, accountability, and transparency need to be clarified in advance. Members of the company need to have

sufficient expertise and a sophisticated understanding of the tools to meet the challenges that the implementation of algorithmic decision-making might face (Barocas and Selbst 2016; Cheng and Hackett 2019; Canhoto and Clear 2020). When using algorithmic decision-making tools, there is an immediate need for transparency and accountability (Tambe et al. 2019). Concerning transparency, this means generating an understanding of how the algorithm operates (e.g., how the algorithm uses data and weighs specific criteria) and disclosing the conditions for the algorithmic decision. Transparency comes along with interpretability and explainability; that is, how the algorithm interacts with the specific data and how it operates in a specific context. Therefore, domain knowledge and knowledge about the programming are indispensable (see Sect. 2.2). Finally, accountability is the acceptance of the responsibility for actions and decisions supported or conducted by algorithms. Companies should clearly define humans responsible for using the algorithmic decision-making tool (Lepri et al. 2018).

Furthermore, HR practitioners must consider the consequences of algorithmic decision-making and be aware that there may be a bias in the training data, because this is often a reflection of existing stereotypes (Mann and O’Neil 2016). As a first step, the company needs to define fairness standards (Canhoto and Clear 2020), because algorithms cannot meet all mathematical and social fairness measures simultaneously. Therefore, the algorithms’ vulnerabilities need to be identified to correct mistakes and improve the algorithms (Lindebaum et al. 2019). Additionally, organizations should write down the exact procedure for the sake of transparency. Companies should also seek to achieve the best quality of input data and continuous update of the used data (Persson 2016). Companies should avoid biased training data (avoiding historical bias) or that certain groups or personal characteristics of interest are underrepresented (avoiding representation bias). Most data sets profit from the renewal of the data to test if the statistical patterns and relationships are still accurate. Notably, in the HRM context, the dynamic nature of personal development needs to be considered, since employees develop and change over time (Simbeck 2019). Thus, it is important to verify and audit the whole process on a regular basis (Kim 2016). Companies should implement a data quality control process to develop quality metrics, collect new data, evaluate data quality, and remove inaccurate data from the training data set. For example, for CV and résumé screening, companies could apply blind hiring, which means removing personally identifiable information on the documents (Yarger et al. 2019; Raghavan et al. 2020).

If the companies use algorithms provided by an external service provider, the algorithms’ code and training data are not transparent for the companies (Raghavan et al. 2020; Sánchez-Monedero et al. 2020). Following the company’s standards mentioned above, HR managers should try to get detailed information about the data sets, the codes, and the procedures and measures of the service provider to prevent biases. Furthermore, HR managers should discuss multiple options that can reduce bias, such as weighing or removing certain indicators that highly correlate with attributes (Yarger et al. 2019).

Due to the lack of intuition and subjective judgment skills when an algorithm decides about a human, employees perceive the decision made by an algorithm as

less fair and trustworthy (Lee 2018). Moreover, pure algorithmic decisions evoke negative feelings (Lee 2018). An implication to prevent anger among the applicants or employees is a disclosure of the nature of the decision made by an algorithm (Cheng and Hackett 2019). A short-term solution to avoid a decrease in the acceptance could be a balanced approach between algorithmic and human decision-making, which means that the algorithm makes a suggestion, but a human checks or even makes the final decision. Hence, algorithmic decision-making seems to be an indispensable tool for assistance in the decision, but human expertise is still necessary (Yarger et al. 2019).

Of course, these practical implications are not limited to HR recruitment and HR development; other HR functions might benefit from these insights, as well. In other HR functions, employees should be informed and, if possible, involved in the algorithms or AI's implementation process. Responsibilities and accountability should be clarified in advance, privacy should be respected, and the possibility for employee voice should be acknowledged. Moreover, they should seek adequate input data and implement data quality checks, which goes along with updating the data regularly. If an external provider is in charge of programming and providing the algorithm, the data and the algorithm should be adapted to the company and should not be adopted without knowing the input data, the conditions for the algorithmic outcomes, and the potential pitfalls of the algorithms.

7 Conclusion

This paper aimed at reviewing current research on algorithmic decision-making in the HRM context, highlighting ethical issues related to algorithms, and outlining implications for future research. The article contributes to a better understanding of the existent research field and summarizes the existing evidence and future research avenues in the highly important topic of algorithmic decision-making. Undoubtedly, the existing studies advanced our understanding of how companies use algorithmic decision-making in HR recruitment and HR development, when, and why unfairness or biases occur in algorithmic decision-making. However, our review suggests that the ongoing debates in computer science on fairness and potential discrimination of algorithms require more attention in leading management journals. Since organizations increasingly implement algorithmic decision tools to minimize human bias, save costs, and automate their processes, our review shows that algorithms are not neutral or free of biases, because a computer has generated a certain decision. Humans should still play a critical and important role in the good governance of algorithmic decision-making.

Acknowledgements We thank Maike Giefers, Hannah Kaiser, and Anna Nieter, and Shirin Riazzy for their support.

Funding Not applicable for that section.

Data availability All material is available upon request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

List of employed keywords

Algorithm

- Algorithm*
- “Algorithmic model*”
- “Data-algorithm*”
- “Algorithmic decision-making”, “algorithmic decision*”
- “Artificial intelligence”
- “Facial expression tool*”, “facial expression processing*”
- “Language processing*”, “natural language processing*”
- “Recommender system*”
- “Search engine*”

Discrimination

- Discrimination*
- Discriminat*
- Classification*, “classification problem*”, “classification scheme*”
- “Algorithmic discrimination*”, “algorithmic bias discrimination*”
- “Preventing discrimination*”
- Anti-discrimination*, non-discrimination*
- Gender, age, sex, sexism, origin
- “Gender-based inequalities”
- “Difference* among demographic group*”
- Ethic*, “ethical implication*”
- “Data mining discrimination*”
- Favoritism, favouritism
- “Unfair treatment*”

Fairness

- Fair*, unfair*
- “Perceived fairness”, “algorithmic fairness”
- “Fairness word*”, “fairness speech*”, “fairness recommendation*”
- Equal*, equit*, inequal*, “equal opportunit*”
- Transparen*
- Legal*, right*
- Truth
- Impartial*
- Correct*
- Justice^a
- Adverse impact^a

Evaluation

- Evaluat*
- Judgement*, “algorithmic judgement*”, “human judgement*”, “mechanical judgement*”
- Rank*
- Rate*
- Measure*
- Valuation*

Bias

- Bias*
- “Algorithmic bias*”, “national bias*”, gender-bias*, “decision-making bias*”, “human bias*”, “technical bias*”
- “Implicit bias* in algorithm*”
- “Dealing with bias*”
- “Pattern distortion*”
- Pre-justice*
- Preconception*
- Tendenc*
- Prone*

Data mining

- Data*
- “Data set*”

HRM

- “Human Resource*”, “Human Resource Management”
- Management
- “Applicant selection*”, “employee selection*”
- “Algorithm-based HR decision-making”
- “Recruitment process*”, “application process*”, “selection process*”
- Recruitment*, online-recruitment*
- “Personnel decision*”, “personnel selection*”
- “People analytic*”, “HR analytic*”
- “Job advertisement*”
- “Online personalization*”

^aRobustness check.

References

- SandAbove. 2020. <https://www.8andabove.com>. Accessed 28 Feb 2020.
- Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through optimization: how Facebook’s ad delivery can lead to skewed outcomes. *arXiv preprint arXiv:1904.02095*.
- Anderson, Neil. 2003. Applicant and recruiter reactions to new technology in selection: a critical review and agenda for future research. *International Journal of Selection and Assessment* 11 (2–3): 121–136.
- Arrow, Kenneth. 1973. The theory of discrimination. *Discrimination in Labor Markets* 3 (10): 3–33.
- Barfield, Woodrow, and Ugo Pagallo. 2018. *Research handbook on the law of artificial intelligence*. Cheltenham: Edward Elgar Publishing.
- Barocas, Solon, and Andrew D. Selbst. 2016. Big data’s disparate impact. *California Law Review* 104: 671.
- Bauer, Talya N., Donald M. Truxillo, Rudolph J. Sanchez, Jane M. Craig, Philip Ferrara, and Michael A. Campion. 2001. Applicant reactions to selection: development of the selection procedural justice scale (SPJS). *Personnel Psychology* 54 (2): 387–419.
- Bengio, Yoshua, Ian Goodfellow, and Aaron Courville. 2017. *Deep learning*. Cambridge: MIT press.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. Implicit discrimination. *American Economic Review* 95 (2): 94–98.
- Bobko, Philip, and C.J. Bartlett. 1978. Subgroup validities: differential definitions and differential prediction. *Journal of Applied Psychology* 63: 12–14.
- Bogen, Miranda. 2019. All the ways hiring algorithms can introduce bias. *Harvard Business Review*, May 6. <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias>.
- Bozdag, Engin. 2013. Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15 (3): 209–227.
- Burdon, Mark, and Paul Harpur. 2014. Re-conceptualising privacy and discrimination in an age of talent analytics. *UNSWLJ* 37:679.
- Burke, Robin, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on fairness, accountability and transparency*. <http://proceedings.mlr.press>.
- Canhoto, Ana Isabel, and Fintan Clear. 2020. Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential. *Business Horizons* 63 (2): 183–193.
- Cappelli, Peter. 2019. Data science can’t fix hiring (yet). *Harvard Business Review* 97 (3): 56–57.
- Cappelli, Peter, Prasanna Tambe, and Valery Yakubovich. 2020. Can data science change human resources? In *The future of management in an AI world*, Berlin: Springer: 93–115.

- Carey, Dennis, and Matt Smith. 2016. How companies are using simulations, competitions, and analytics to hire. *Harvard Business Review*. <https://hbr.org/2016/04/how-companies-are-using-simulations-competitions-and-analytics-to-hire>.
- Cascio, Wayne F., and Herman Aguinis. 2013. *Applied psychology in human resource management*. London: Pearson Education.
- Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. Productivity and selection of human capital with machine learning. *American Economic Review* 106 (5): 124–127.
- Chamorro-Premuzic, Tomas, Dave Winsborough, Ryne A. Sherman, and Robert Hogan. 2016. New talent signals: shiny new objects or a brave new world? *Industrial and Organizational Psychology* 9 (3): 621–640.
- Chamorro-Premuzic, Tomas, Reece Akhtar, Dave Winsborough, Ryne A Sherman. 2017. The datafication of talent: how technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences* 18:13–16.
- Chander, Anupam. 2016. The racist algorithm. *Michigan Law Review* 115: 1023.
- Chen, Le, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the impact of gender on rank in resume search engines. In *Proceedings of the 2018 chi conference on human factors in computing systems*: 1–14. <https://doi.org/10.1016/j.hrmr.2019.100698>.
- Cheng, Maggie M., and Rick D. Hackett. 2019. A critical review of algorithms in HRM: definition, theory, and practice. *Human Resource Management Review* 100698.
- Citron, Danielle Keats, and Frank Pasquale. 2014. The scored society: due process for automated predictions. *Washington Law Review* 89: 1.
- Cohen-Charash, Yochi, and Paul E. Spector. 2001. The role of justice in organizations: a meta-analysis. *Organizational Behavior and Human Decision Processes* 86 (2): 278–321.
- Cropanzano, Russell, David E. Bowen, and Stephen W. Gilliland. 2007. The management of organizational justice. *Academy of Management Perspectives* 21 (4): 34–48.
- Crossan, Mary M., and Marina Apaydin. 2010. A multi-dimensional framework of organizational innovation: a systematic review of the literature. *Journal of Management Studies* 47 (6): 1154–1191.
- Danks, David, and Alex John London. 2017. Algorithmic bias in autonomous systems. In *IJCAI*: 4691–4697.
- Dastin, Jeffrey. 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. San Francisco: Reuters.
- Datta, Amit, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies* 2015 (1): 92–112.
- Daugherty, Paul R., and H.J. Wilson. 2018. *Human+ machine: reimagining work in the age of AI*. Boston: Harvard Business Press.
- Deloitte. 2018. Mensch bleibt Mensch - auch mit algorithmen im recruiting. Wo der Einsatz von Algorithmen hilfreich ist und wo nicht. <https://www2.deloitte.com/de/de/pages/careers/articles/algorithmen-im-recruiting-prozess.html>. Accessed 12 Sept 2019.
- Deloitte. 2020. State of AI in the enterprise – 3rd edition results of the survey of 200 AI experts on artificial intelligence in German companies. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/technology-media-telecommunications/DELO-6418_State%20of%20AI%202020_KS4.pdf. Accessed 10 Jun 2020.
- Deng, Li., and Yu. Dong. 2014. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing* 7 (3–4): 197–387.
- Diakopoulos, Nicholas. 2015. Algorithmic accountability: journalistic investigation of computational power structures. *Digital Journalism* 3 (3): 398–415.
- Dreisbach, Caitlin, Theresa A. Koleck, Philip E. Bourne, Suzanne Bakken. 2019. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International Journal of Medical Informatics* 125:37–46.
- Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*: ACM: 214–226.
- Ferguson, Christopher J., and Michael T. Brannick. 2012. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods* 17 (1): 120.

- Florentine, S. 2016. How artificial intelligence can eliminate bias in hiring. *CIO Magazine*. <https://www.cio.com/article/3152798/artificial-intelligence-how-artificial-intelligence-can-eliminate-bias-in-hiring.html>. Accessed 03 Mar 2020.
- Friedman, Batya, and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on Information Systems* 14 (3): 330–347.
- Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Huldgren. 2013. Value sensitive design and information systems. In *Early engagement and new technologies: opening up the laboratory*, Dodrecht: Springer: 27–55.
- Frijters, Paul. 1998. Discrimination and job-uncertainty. *Journal of Economic Behavior & Organization* 36 (4): 433–446.
- Gil-Lafuente, Anna María, and Young Kyun Oh. 2012. Decision making to manage the optimal selection of personnel in the hotel company applying the hungarian algorithm. *The International Journal of Management Science and Information Technology* 6-(Oct-Dec): 27–42.
- Gilliland, Stephen W. 1993. The perceived fairness of selection systems: an organizational justice perspective. *Academy of Management Review* 18 (4): 694–734.
- Goodfellow, Ian, Y. Bengio, and A. Courville. 2016. Machine learning basics. *Deep Learning* 1: 98–164.
- Gough, David, Sandy Oliver, and James Thomas. 2017. *An introduction to systematic reviews*. London: Sage.
- Guchait, Priyanko, Tanya Ruetzler, Jim Taylor, and Nicole Toldi. 2014. Video interviewing: a potential selection tool for hospitality managers—a study to understand applicant perspective. *International Journal of Hospitality Management* 36: 90–100.
- Hardt, Moritz, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*: 3315–3323.
- Hausknecht, John P., David V. Day, and Scott C. Thomas. 2004. Applicant reactions to selection procedures: an updated model and meta-analysis. *Personnel Psychology* 57 (3): 639–683.
- HireVue. 2019. <https://www.hirevue.com>. Accessed 01.Jan 2020.
- Hiemstra, Annemarie MF., Janneke K. Oostrom, Eva Deros, Alec W. Serlie, and Marise Ph Born. 2019. Applicant perceptions of initial job candidate screening with asynchronous job interviews: does personality matter? *Journal of Personnel Psychology* 18 (3): 138.
- Hoffmann, Anna Lauren. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22 (7): 900–915.
- Horton, John J. 2017. The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics* 35 (2): 345–385.
- Huselid, Mark A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal* 38 (3): 635–672.
- IBM. 2020. IBM Watson Career Coach for career management. <https://www.ibm.com/talent-management/career-coach>. Accessed 20 Apr 2020.
- Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. 1996. Reinforcement learning: a survey. *Journal of Artificial Intelligence Research* 4: 237–285.
- Kahneman, Daniel, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
- Kaibel, Chris, Irmela Koch-Bayram, Torsten Biemann, and Max Mühlenbock. 2019. Applicant perceptions of hiring algorithms-uniqueness and discrimination experiences as moderators. In *Academy of Management Proceedings*: Academy of Management Briarcliff Manor, NY 10510.
- Kaplan, Andreas, and Michael Haenlein. 2019. Siri, Siri, in my hand: who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons* 62 (1): 15–25.
- Kauermann, Goeran, and Helmut Kuechenhoff. 2010. *Stichproben: Methoden und praktische Umsetzung mit R*. Berlin: Springer.
- Kellogg, Katherine C., Melissa A. Valentine, Angéle Christin. 2020. Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals* 14(1):366–410.
- Kim, Pauline T. 2016. Data-driven discrimination at work. *William & Mary Law Review* 58: 857.
- Kim, P. T. 2017. Data-Driven Discrimination at Work. *William & Mary Law Review*, 58(3):857.
- Kim, Pauline T., and Sharion Scott. 2018. Discrimination in online employment recruiting. *Louis ULJ* 63: 93.
- Kuncel, Nathan R., David M. Klieger, Brian S. Connelly, and Deniz S. Ones. 2013. Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. *Journal of Applied Psychology* 98 (6): 1060.

- Lambrecht, Anja, and Catherine Tucker. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science* 65 (7): 2966–2981.
- Langer, Markus, Cornelius J. König, and Andromachi Filiti. 2018. Information as a double-edged sword: the role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior* 81: 19–30. <https://doi.org/10.1016/j.chb.2017.11.036>.
- Langer, Markus, Cornelius J. König, and Maria Papathanasiou. 2019. Highly automated job interviews: acceptance under the influence of stakes. *International Journal of Selection and Assessment*. <https://doi.org/10.1111/ijasa.12246>.
- Leclercq-Vandelannoitte, Aurélie. 2017. An Ethical Perspective on Emerging Forms of Ubiquitous IT-Based Control. *Journal of Business Ethics* 142 (1):139–154.
- Lee, Min Kyung. 2018. Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5 (1): 2053951718756684.
- Lee, Min Kyung, and Su Baykal. 2017. Algorithmic mediation in group decisions: fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*: ACM: 1035–1048.
- Lee, In., and Yong Jae Shin. 2020. Machine learning for enterprises: applications, algorithm selection, and challenges. *Business Horizons* 63 (2): 157–170.
- Leicht-Deobald, Ulrich, Thorsten Busch, Christoph Schank, Antoinette Weibel, Simon Schafheitle, Isabelle Wildhaber, and Gabriel Kasper. 2019. The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics* 160 (2): 377–392.
- Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* 31 (4): 611–627.
- Leventhal, Gerald S. 1980. What should be done with equity theory? In *Social exchange*, New York: Springer: 27–55.
- Lindebaum, Dirk, Mikko Vesa, and Frank den Hond. 2019. Insights from the machine stops to better understand rational assumptions in algorithmic decision-making and its implications for organizations. *Academy of Management Review*. <https://doi.org/10.5465/amr.2018.0181>.
- Lipsey, Mark W., and David B. Wilson. 2001. *Practical meta-analysis*. Thousand Oaks: SAGE publications Inc.
- Mann, Gideon, and Cathy O’Neil. 2016. Hiring algorithms are not neutral. *Harvard Business Review* 9. <https://hbr.org/2016/12/hiring-algorithms-are-not-neutral>.
- McCarthy, Julie M., Talya N. Bauer, Donald M. Truxillo, Neil R. Anderson, Ana Cristina Costa, and Sara M. Ahmed. 2017. Applicant perspectives during selection: a review addressing “So what?”, “What’s new?”, and “Where to next?” *Journal of Management* 43 (6): 1693–1725.
- McColl, Rod, and Marco Michelotti. 2019. Sorry, could you repeat the question? Exploring video-interview recruitment practice in HRM. *Human Resource Management Journal* 29 (4): 637–656.
- McDonald, Kathleen, Sandra Fisher, and Catherine E. Connelly. 2017. e-HRM systems in support of “smart” workforce management: an exploratory case study of system success. *Electronic HRM in the Smart Era* 87–108. <https://doi.org/10.1108/978-1-78714-315-920161004>
- Meade, Adam W., and Michael Fetzner. 2009. Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods* 12 (4): 738–761.
- Miller 2015. Can an algorithm hire better than a human. The New York Times. <https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html>. Accessed 13 sep 2019.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* 151 (4): 264–269.
- Möhlmann, M., and L. Zalmanson. 2017. Hands on the wheel: navigating algorithmic management and Uber drivers’. In *Autonomy’, in proceedings of the international conference on information systems (ICIS)*, Seoul South Korea: 1–17.
- Morrison, Andra, Julie Polisena, Don Husereau, Kristen Moulton, Michelle Clark, Michelle Fiander, Monika Mierzwinski-Urban, Tammy Clifford, Brian Hutton, and Danielle Rabb. 2012. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *International Journal of Technology Assessment in Health Care* 28 (2): 138–144.
- Murphy, Kevin P. 2012. *Machine learning: a probabilistic perspective*. Cambridge: MIT press.

- Naim, Iftekhar, Md Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque. 2016. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing* 9 (2): 191–204.
- Ötting, Sonja K., and Günter. W. Maier. 2018. The importance of procedural justice in human–machine interactions: intelligent systems as new decision agents in organizations. *Computers in Human Behavior* 89: 27–39.
- Paschen, Ulrich, Christine Pitt, and Jan Kietzmann. 2020. Artificial intelligence: Building blocks and an innovation typology. *Business Horizons* 63 (2): 147–155.
- Pasquale, Frank. 2015. *The black box society*. Cambridge: Harvard University Press.
- Persson, Anders. 2016. Implicit bias in predictive data profiling within recruitments. In *IFIP International Summer School on Privacy and Identity Management*. Springer.
- Petticrew, Mark, and Helen Roberts. 2008. *Systematic reviews in the social sciences: a practical guide*. Hoboken: John Wiley & Son.
- Podsakoff, Philip M., Scott B. MacKenzie, Daniel G. Bachrach, and Nathan P. Podsakoff. 2005. The influence of management journals in the 1980s and 1990s. *Strategic Management Journal* 26 (5): 473–488.
- Prassl, Jeremias. 2018. *Humans as a service: the promise and perils of work in the gig economy*. Oxford: Oxford University Press.
- Precire. 2020. Precire technologies. <https://precire.com/>. Accessed 03 Jan 2020.
- Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*.
- Roscher, Ribana, Bastian Bohn, Marco F. Duarte, and Jochen Garcke. 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8: 42200–42216.
- Rosenblat, Alex, Tamara Kneese, and Danah Boyd. 2014. Networked employment discrimination. Open Society Foundations' Future of Work Commissioned Research Papers.
- Rosenblat, Alex, and Luke Stark. 2016. Algorithmic labor and information asymmetries: a case study of Uber's drivers. *International Journal of Communication* 10: 27.
- Roth, Philip L., Huy Le, Oh. In-Sue, Chad H. Van Iddekinge, and Steven B. Robbins. 2017. Who ru?: On the (in) accuracy of incumbent-based estimates of range restriction in criterion-related and differential validity research. *Journal of Applied Psychology* 102 (5): 802.
- Russell, Stuart J., and Peter Norvig. 2016. *Artificial intelligence: a modern approach*. London: Pearson Education Limited.
- Ryan, Ann Marie, and Robert E. Ployhart. 2000. Applicants' perceptions of selection procedures and decisions: a critical review and agenda for the future. *Journal of Management* 26 (3): 565–606.
- Sajjadi, Sima, Aaron J. Sojourner, John D. Kammeyer-Mueller, and Elton Mykerezzi. 2019. Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000405>.
- Sánchez-Monedero, Javier, Lina Dencik, and Lilian Edwards. 2020. 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*: 458–468.
- Savage, David, and Richard A. Bales. 2017. Video games in job interviews: using algorithms to minimize discrimination and unconscious bias. *ABA Journal of Labor & Employment Law* 32.
- Siddaway, Andy P., Alex M. Wood, and Larry V. Hedges. 2019. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology* 70: 747–770.
- Silverman, Rachel Emma, and Nikki Waller. 2015. The algorithm that tells the boss who might quit. *Wall Street Journal*. <http://www.wsj.com/articles/the-algorithm-that-tells-the-boss-who-might-quit-1426287935>.
- Simbeck, K. 2019. HR analytics and ethics. *IBM Journal of Research and Development* 63 (4/5): 1–9.
- Stone, Diana L. Dadrack, Kimberly M. Lukaszewski, Richard Johnson. 2015. The influence of technology on the future of human resource management. *Human Resource Management Review* 25 (2): 216–231.
- Suen, Hung-Yue., Mavis Yi-Ching. Chen, and Lu. Shih-Hao. 2019. Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior* 98: 93–101.

- Sumser, John. 2017. Artificial intelligence: ethics, liability, ownership and HR. *Workforce Solutions Review* 8 (3): 24–26.
- Suresh, Harini, and John V. Gutttag. 2019. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*.
- Tambe, Prasanna, Peter Cappelli, and Valery Yakubovich. 2019. Artificial intelligence in human resources management: challenges and a path forward. *California Management Review* 61 (4): 15–42.
- van Esch, Patrick, J. Stewart Black, and Joseph Ferolie. 2019. Marketing AI recruitment: the next phase in job application and selection. *Computers in Human Behavior* 90: 215–222.
- Van Hoye, G. 2014. Word of mouth as a recruitment source: an integrative model. In Yu, K.Y.T. and Cable, D.M. (eds), *The Oxford Handbook of Recruitment*. Oxford: Oxford University Press: 251–268.
- Varghese, Jacob S., James C. Moore, and Andrew B. Whinston. 1988. Artificial intelligence and the management science practitioner: rational choice and artificial intelligence. *Interfaces* 18 (4): 24–35.
- Vasconcelos, Marisa, Carlos Cardonha, and Bernardo Gonçalves. 2018. 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.
- Veale, Michael, and Reuben Binns. 2017. Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data & Society* 4 (2): 2053951717743530.
- Walker, Joseph. 2012. Meet the new boss: big data. *Wall Street Journal*. <https://online.wsj.com/article/SB10000872396390443890304578006252019616768.html>. Accessed 13 Mar 2020
- Williams, Betsy Anne, Catherine F Brooks, Yotam Shmargad. 2018. How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications. *Journal of Information Policy* 8:78–115.
- Wolpert, David H., and William G. Macready. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1 (1): 67–82.
- Woodruff, Allison, Sarah E Fox, Steven Rouso-Schindler, and Jeffrey Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
- Woods, Stephen A., Sara Ahmed, Ioannis Nikolaou, Ana Cristina Costa, and Neil R. Anderson. 2020. Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology* 29 (1): 64–77.
- Yarger, Lynette, Fay Cobb Payton, and Bikalpa Neupane. 2019. Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*. <https://doi.org/10.1108/OIR-10-2018-033>. Accessed 3 Mar 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.