Is artificial intelligence (AI) research biased and conceptually vague? A systematic review of research on bias and discrimination in the context of using AI in human resource management

Ivan Kekez [a,*] , Lode Lauwaert [b] , Nina Begičević Ređep [a]

[a] *University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000, Varaždin, Croatia*
[b] *Institute of Philosophy, KU Leuven, Kardinaal Mercierplein 2, 3000, Leuven, Belgium*

ABSTRACT

This paper presents a systematic review of 64 papers using the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) of research on bias and discrimination in the context of using Artificial Intelligence (AI). Specifically, while limiting the scope to research in HRM, it aims to answer three questions that are relevant to the research community. The first question is whether research papers define the terms 'bias' and 'discrimination', and if so how. Second, given that there are different forms of bias and discrimination, the question is exactly which ones are being investigated. Are there any forms of bias and discrimination that are underrepresented? The third question is whether a negativity bias exists in research on bias and discrimination in the context of AI. The answers to the first two questions point to some research problems. The review shows that in a substantial number of papers, the terms 'bias' and 'discrimination' are not or hardly defined. Furthermore, there is a disproportionate focus among researchers on bias and discrimination related to skin tone (racism) and gender (sexism). In the discussion, we provide reasons why this is undesirable for both scientific and extratheoretical reasons. The answer to the last question is negative. There is a relatively good balance between research that zooms in on the positive effects of AI on bias and discrimination, and research that deals with AI leading to (more) bias and discrimination.

1. Introduction

It is a truism that AI has penetrated (almost) every domain of society within the short span of just over a decade, from education and research over banking and politics to sports and the legal system. In light of this, one might expect that research on AI has also increased, which then should result in more publications and participants at conferences on AI. This expectation is supported by data. In 2012 – the year Geoffrey Hinton presented the neural network-based face recognition system AlexNet, which boosted AI's commercial breakthrough – about 196,000 articles on AI were published by scholars in both English and Chinese, ranging from journal articles to conference papers to books. Ten years later, that number had more than doubled. The number of people attending conferences on AI in recent years is in line with this. Despite the COVID-19 pandemic, the number of attendees would have roughly increased more than fivefold over the past decade (Roser).

The same can be said of research on ethical topics related to AI.

Research, both by philosophers and non-philosophers, is also increasing on the ethical questions about AI. At least that is one of the conclusions of the Stanford University *Artificial Intelligence Index Report 2022*, which annually provides an overview of trends in AI research: "Research on fairness and transparency in AI has exploded since 2014, with a fivefold increase in related publications at ethics-related conferences. Algorithmic fairness and bias have shifted from being primarily an academic pursuit to becoming firmly embedded as a mainstream research topic with wide-ranging implications. Researchers with industry affiliations contributed 71% more publications year over year at ethics-focused conferences in recent years." (Zhang et al., 2022)

There is a wide variety of topics being written about by both philosophers and non-philosophers (including, for example, legal scholars and psychologists) in the papers in which they focus on ethics in the context of AI. These themes, among other things, include responsibility, job replacement, disinformation, privacy, bias, and discrimination. It is especially the last two topics that frequently reoccur in HRM (Storm

et al., 2023), (Triana et al., 2021). When one reads a paper that delves into ethics and AI, there is a real to very high probability that it will (also, and perhaps even primarily) discuss the risk that the output of a chatbot, for instance, reproduces sexist or racist stereotypes (bias), or that the paper will address the problem of an AI system treating people of color unequally based on their color (discrimination). In short, whether approached by scholars in philosophy or outside of the context of philosophy, when it comes to the ethics of AI, it often concerns bias and discrimination.

The observation that both bias and discrimination are two prominent topics in the research of both philosophers and non-philosophers leads to (at least) three research questions. All of them probe the diversity in research, meaning they aim to determine whether certain concepts or themes are sufficiently present in studies on bias and discrimination in the context of AI.

First, there is the question of whether scholars define bias and discrimination in their papers, and if so, what exact definitions they provide for both terms. The reason we bring this question to the fore is that many terms (including words with an ethical connotation) often do not have a clear-cut meaning, and thus, that they can express different concepts. Consequently, from a scientific point of view, it is undesirable when a paper does not define 'bias' or 'discrimination'. After all, it can lead to it being unclear what exactly the study is and is not about, and it becomes difficult or even impossible to compare the results of different studies, i.e., it is than far from obvious to determine whether the studies replicate or contradict each other. Now, when a systematic review reveals that there is a tendency not to define research terms, this may be a sign for the research community that there is a systematic research deficit that needs to be eliminated. On the other hand, when the terms are actually defined, it is relevant for researchers who focus on bias and discrimination in AI to learn in the review what exactly the definitions are. Indeed, in this way, one can get a picture of certain interpretations of bias and discrimination that are not or insufficiently represented, or scholars can then accurately determine, based on the review, whether new or forthcoming research does or does not align with existing research results.

The second question concerns the characteristic on the basis of which bias or discrimination occurs. After all, both phenomena never arise in general but always stem from a specific characteristic of a particular individual or group. Moreover, there is a wide variety of characteristics on the basis of which someone can be biased or discriminated against. Skin tone and gender are perhaps the two best-known characteristics, but other characteristics such as age, mental or physical disabilities, religion, or physical attractiveness can also be the basis for bias or discrimination. The second research question, therefore, is the following: what forms of bias and discrimination are being studied in the context of AI? Here, too, the concern of unequal distribution plays a role. It is possible that certain forms of bias and discrimination are researched more or less frequently, which is undesirable from both a scientific and moral standpoint. If the systematic review indicates this, this would provide a reason for researchers to address the underrepresentation and overrepresentation.

Third, do we face a negativity bias in research on bias and discrimination in the context of AI? More specifically, can one detect substantially more research on how AI can cause bias and discrimination (undesirable) than on how AI can be useful to address problems with bias and discrimination (desirable)? This question arises for several reasons. First, it is a well-documented scientific fact that humans have a negativity bias (among many other biases), meaning that they are more inclined to focus on bad news rather than good news (Rozin & Royzman, 2001), (Vaish et al., 2008). Additionally, scholars have previously suggested that there is indeed more attention given to the (ethical or not) drawbacks than the benefits of AI (Jecker & Nakazawa, 2022). Investigating whether this is actually the case is relevant. Indeed, if it turns out that there actually is more focus on the problems of AI in terms of bias and discrimination, then this can skew research priorities and outcomes,

and may lead to an unbalanced understanding of AI's capabilities and limitations. By consequence, the review could be a reason for researchers to fill this gap in the literature and to make sure that future provides a more balanced and constructive perspective on AI's role in society.

However, it is practically impossible to collect, screen, and analyze all papers on these three questions. It is thus essential to clearly define the domain and limit the number of papers. Therefore, we have chosen to exclusively focus on papers that discuss bias, discrimination, and AI in the context of Human Resources Management (HRM). There are at least three reasons why we decided to limit our review to HRM specifically. First, the expertise of two of the three authors lies within the research domain of HRM. Second, perhaps the most well-known case commonly used to discuss AI and ethics, particularly bias and discrimination, is precisely about HRM. We refer here to the case of Amazon experimenting with AI to screen job applications for new positions at the American company. This case became widely known because the AI system systematically excluded applications from female candidates, which is clearly a case of discrimination and can be described as 'algo-sexism' (Dastin). The third reason is that while AI solutions can enhance and improve human resource management (HRM) (Murugesan et al., 2023), there are also notable instances of bias and discrimination when AI is applied in this field. Given that HRM decisions directly impact individuals, this makes the issue particularly sensitive and important.

The aim of this paper is to present a systematic review of the scientific literature that addresses these three research questions.

RQ1. Do authors define bias and discrimination in research studies on AI adoption in HRM, and if so, how?

RQ2. What are the main topics discussed in the research studies regarding bias and discrimination in HRM?

RQ3. Do researchers focus more on desirable outcomes than on the undesirable effects of the use of AI?

We proceed as follows. In the following section, we present the research methodology. After that, we provide the results of the study, which are discussed in the subsequent section. We close the paper with a final conclusion.

2. Research methodology

In this section, we provide a systematic literature overview using the Preferred Reporting Items for Systematic reviews and Meta-Analysis, later in text PRISMA framework. The PRISMA framework was originally published in 2009, later to be updated and replaced by the PRISMA 2020 and published in 2021 (Page et al., 2021). In contrast to conventional literature reviews, the PRISMA framework follows a strict protocol designed using guidelines, checklist, explanation boxes and flow diagram. The PRISMA framework prioritizes defining exclusion and inclusion criteria to ascertain the suitability of literature for further analysis or to disprove its relevance (Xames & Topcu, 2024).

The review involves four phases following the PRISMA guidelines: (i) identification, (ii) screening, (iii) eligibility, and (iv) inclusion. Each phase is explained in detail below and broad illustration is shown in Fig. 1.

2.1. Identification phase

The identification phase focused on the selection of electronic scientific databases, as well as the retrieval of literature data need for analysis. The databases included were Web of Science Core Collection (CC), Scopus, IEEE Xplore Digital Library, Science Direct, and ProQuest. The search was performed iteratively through several testing phases in November and December 2023, while content analysis occurred in the first half of 2024. Although the first AI systems were easily interpretable, the growing commercial and empirical success led to the requirement

Fig. 1. PRISMA flow diagram, adopted by Page and colleagues (Page et al., 2021).

for the interpretation of associated terms (Barredo Arrieta et al., 2020), (Dwivedi et al., 2021). With machine learning models increasingly utilized in predictive scenarios (Barredo Arrieta et al., 2020), recent research on AI taxonomy highlights machine learning and deep learning models as the central domains of AI (Graziani et al., 2023), (Schwalbe &

Finzel, 2023). In simple terms, machine learning is a field of AI that describes the system's capacity to learn from data to automate and operationalize the process of analytical model building and task accomplishment. Deep learning is a sophisticated machine-learning concept based on artificial neural networks, prevalent in machine

learning due to its high capability outputs and minimum requirement for human interventions (Janiesch et al., 2021), (Du et al., 2016). Thus, considering the widespread use of machine learning and deep learning models in technical and social domains (Graziani et al., 2023), we decided to form search queries around AI and its subsets: machine learning and deep learning. The broad representation of search queries used to identify literature was framed by the topic theme it represents: (i) „artificial intelligen*" OR „machine learning" OR „deep learning" AND (ii) „discrimination" OR „bias" AND (iii) „human resource*". The total number of records was 317 across five databases. Table 1 displays the database search queries alongside the number of records for each database per year.

All records were extracted in either.csv or.txt format, depending on the export options provided by the respective databases. The selected categories for export included authors' names, article titles, document types, publication years, and abstracts. Additionally, a new category, „*database source*," was added to the spreadsheet, using the following abbreviations for input: WoS (Web of Science), SCP (Scopus), SD (Science Direct), IEEE (IEEE Xplore Digital Library), and ProQ (ProQuest). Considering databases use different abbreviations or descriptions for document types, we have implemented a categorization system that ensures consistency across all exported records (Table 2).

Among the 317 records, 113 were identified as duplicates, leaving 204 records available for further analysis. The records were filtered using the inclusion/exclusion criteria shown in Table 3 below.

We limited our analysis to records categorized as journal articles, conference papers, and review papers, as these publications often undergo a rigorous peer-review process. Journal articles typically receive more thorough review proces than conference papers. However, conference papers provide opportunities to present and discuss preliminary research results and are occasionally upgraded to meet journal article standards (Gonzá et al., 2011). Records categorized as books, book chapters, working papers, dissertations, and other types were excluded from further analysis. The distribution of included records showed 107 journal articles, 42 conference papers, and 9 reviews. Records listed in languages other than English were excluded. Due to the relatively low number of included records, we decided to include records spanning the entire time range without setting restrictions based on the publication date. The total number of excluded records based on the given criteria was 46.

2.2. Screening phase

The next stage of the PRISMA framework involved a review process for included records based on title and abstract review. The prior phase solely focused on filtering exported records using straightforward mechanics such as database searches using queries, data export, minimal data pre-processing in Microsoft Excel, and defining and implementing inclusion and exclusion criteria. As we transition from 'data-driven' information, solely based on extracted categories provided by databases, to content analysis of the extracted material, we will subsequently refer

Table 2
Records categorization based on the initially listed document types.

Document type	Originally listed document type in Web of Science, Scopus, Science Direct, IEEE Xplore Digital Library, and ProQuest
Journal Article	Article, Article In Press, Article In Early Access, IEEE Journals, Peer Review Journal Article, Journal Article
Conference Paper	Conference paper, Conference Proceedings, IEEE Conferences, Proceedings Paper
Review	Review
Book/Book Chapter	Book, Book Chapter, Wiley IEEE Press eBook Chapter
Working Paper	Working paper
Dissertation	Dissertation Abstract/Dissertation, Dissertation Thesis
Other	Editorial, Feature, News, Periodical

Table 3
Inclusion and exclusion criteria followed by the number of included and excluded records per document type.

Criteria	Inclusion	Exclusion
Document type	Journal Article (#107), Conference Paper (#42), Review (#9)	Book/Book Chapter (#11), Working Paper (#8), Dissertation (#17), Other (#10)
Language	English	Other languages
Publication Date Range	All	–
Total	158 records	46 records

to the exported results as publications categorized as journal articles and conference papers. Given the importance of the screening phase, it is essential to establish a set of rules or questions that assist the screener in identifying eligible publications for further analysis (Polanin et al., 2019). Additionally, because abstracts should provide a reader with the subject background, methods, results, and conclusion in a limited number of words, there is a possibility of misinterpretation due to insufficient information, poor quality, or deception using unjustified assumptions, leading readers to form a biased understanding of the paper (Andrade, 2011). For that matter, we formed a short set of guidelines, reasoning the publication exclusion or inclusion for in-depth analysis (Table 4). Furthermore, it is worth mentioning that during this phase, we focused on the research approach, themes discussed, and keywords rather than the results of the publications or the conclusions addressed in the abstract.

At the end of the abstract screening process, 83 publications were recognized as relevant and set for download. In cases where the download link is not directly available in the database, the search is redirected to the publication source or the private profiles of the authors (e.g., ResearchGate). Four publications have limited access (*marked as italic in the Table below*). Source titles with links and the number of downloads are shown in Table 5.

Table 1
Representation of scientific databases, corresponding database queries and the number of records per year.

Database	Query design	All time	2024	2023	2022	2021	2020	2019	>2019
Web of Science CC	TS = (Artificial Intelligen* OR Machine Learning OR Deep Learning) AND TS = (Discrimination OR Bias) AND TS = (Human NEAR/5 Resource*)	74	1	15	19	21	8	6	4
Scopus	TITLE-ABS-KEY ("artificial intelligen*" OR "machine learning" OR "deep learning") AND TITLE-ABS-KEY ("discrimination" OR "bias") AND TITLE-ABS-KEY ("human resource*")	113	–	28	32	26	10	8	9
Science Direct	Title, abstract, keywords: (artificial intelligence OR machine learning OR deep learning) AND (discrimination OR bias) AND (human resource)	23	–	6	4	7	2	1	3
IEEE Xplore Digital Library	("Abstract":artificial intelligen* OR machine learning OR deep learning) AND ("Abstract": discrimination OR bias) AND ("Abstract":human NEAR/5 resource*)	14	–	2	3	3	1	1	4
ProQuest	abstract(artificial intelligen* OR machine learning OR deep learning) AND abstract (discrimination OR bias) AND abstract(human NEAR/5 resource*)	93	–	19	30	12	9	11	12

Table 4

The abstract screening categories and guidelines describing inclusion and exclusion criteria.

Reasoning for exclusion	Keynotes from abstract – included publications	A close decision *
Research keywords: AI, ML, DL, bias, discrimination, HRM, etc. No [research keywords] methods nor techniques are discussed nor mentioned in the abstract. The abstract considers [research keywords] approaches/themes but shifts the focus towards another direction. The abstract considers [research keyword] as a part of comparative analysis between different approaches rather than its practical usage	Short keynotes or paragraph from the abstract that justify publication inclusion for further analysis Example: **AI**, **recruitment**, **applicants**, **bias** „*While **AI** can expedite the **recruitment process**, evidence from the industry, however, shows that AI-recruitment systems (AIRS) may fail to achieve **unbiased decisions** about **applicants**. There are risks of encoding **biases** in the **datasets and algorithms of AI** which lead AIRS to replicate and amplify **human biases**.*" (Soleimani, Intezari, & Pauleen, 2021)	* - The paper is included, but a more detailed examination is required during the Eligibility phase.

Table 5

Publication source and number of downloads.

Source		Number of publications
Open access - link for download provided in scientific databases using database proxy	Web of Science, Scopus, IEEE Xplore, Science Direct, ProQuest	**53**
National and University Library in Zagreb, Croatia	https://nsk.hr/en/	**15**
European Agency for Safety and Health at Work	https://osha.europa.eu/en	**1**
Google Scholar	https://scholar.google.com/	**1**
RecSys Conference	https://recsys.acm.org/	**1**
Research Gate	https://www.researchgate.net/	**4**
Sage Journals/Publications	https://us.sagepub.com/en-us/nam/home	**2**
Scholar Space	https://scholarspace.manoa.hawaii.edu/home	**1**
Semantic Scholar	https://www.semanticscholar.org/	**1**
Limited Access	-	*(4)*
TOTAL		**79**

2.3. Eligibility phase

The eligibility phase marks the third stage in determining the publication relevance for content analysis in the final phase. For abstracts that passed the screening phase and had accessible full versions for download, a comprehensive review is necessary. In this phase, any publications that deviate from the research goal or may been overlooked based on technical criteria are excluded from further analysis. By leveraging our research goal during the publication review, we have identified 24 publications that are not in alignment with our research objective or did not fulfill the required technical criteria. The majority of the excluded publications were irrelevant to the topic, with a few not meeting the minimal technical criteria. In addition, we expanded the number of publications by adding nine publications from additional sources. These publications were not identified using database search

queries but were identified through a manual search on Google Scholar using predefined keywords. They were accessible for download and aligned with the research objective outlined in the introduction. Additionally, all technical inclusion criteria for these publications were met. This approach aimed to partially mitigate the limitation of potentially overlooking relevant studies indexed in other sources. As a result, the final analysis included 64 publications.

2.4. Inclusion phase

For the textual analysis, we chose to perform an inductive content analysis. Inductive content analysis involves analyzing chunks of textual data and setting coding rules to shrink the big-picture perspective into more manageable categories. That includes reading the full text and finding essential points related to the research questions (Vears & Gillam, 2022). The unit of analysis was the whole publication included in the final phase. The coding template includes 9 categories that summarize information from the included publications. In short, this involves information on publication data (authors, title, year of publication), research focus, details on bias and discrimination (definitions, types and topics), and authors' perspectives on AI implications in practice. Finally, a structure and description of the coding template for content analysis are listed in Table 6.

3. Results

This section discusses the results of the finally included publications in the content analysis. A total of 64 publications met the inclusion criteria and were included in the final analysis for the systematic literature review. The analysis aims to identify thematic overlaps and associations between parts of the included publications. The first part provides a descriptive analysis of publications included in the content analysis, followed by the second part, a numeric breakdown of the research questions mentioned in the previous section, represented through the subsections. Lastly, in the third part, we have summarized the research contributions, providing readers with a consolidated overview.

Table 6

A coding template used to outline the categories, the corresponding values, and detailed descriptions for publication analysis.

Category	Value	Description
Authors	Text	Name and surname of listed authors
Publication title	Text	Title of the included publication in the content analysis
Publication year	Integer	A numeric field indicating the year of publication
Research focus	Text	Focus (goal) of the publication included in the content analysis
Bias and/or discrimination	B = bias; D = discrimination; BD = bias and discrimination	Does the publication discuss bias, discrimination, or both terms?
Bias and discrimination definitions	Text	Definitions on discussed terms
Bias and discrimination topics	Text	What are the main topics identified regarding bias and discrimination? Broadly include age, gender, ethnicity, race, religious beliefs, disabilities, physical attributes, etc.
Positive implications	Text = Discussed Blank = Not discussed in detail	Researchers' discussion on the positive and negative implications of AI adoption in the context of HRM
Negative implications	Text = Discussed Blank = Not discussed in detail	

3.1. Descriptive analysis

Out of the 64 publications included for content analysis, 17 were conference papers and 47 were journal articles. The earliest publication dates back to 2015, indexed in Scopus, followed by a conference paper listed in 2017 and indexed in Web of Science, Scopus, IEEE Explore Digital Library, and ProQuest, and another conference paper listed in 2018 and indexed in Scopus. The time span of the included publications ends with one journal article in early access, scheduled for publication in 2024. A graph containing information on identified and extracted papers on December 8th, 2023., is shown below (see Figs. 2 and 3).

The distribution of the included publications (Fig. 4), based on the database source, shows that the highest number of publications were indexed in Scopus alone (18), followed by publications indexed in Web of Science, Scopus, and ProQuest (11), as well as the intersection between Web of Science and Scopus (11).

All conference papers (17) were indexed in Scopus. The highest number of conference papers was indexed in Scopus alone (12), followed by the intersection of Web of Science, Scopus, IEEE Digital Library, and ProQuest (3). Two conference papers were indexed in two separate groups: one combining Web of Science and Scopus and the other combining Scopus, IEEE Digital Library, and ProQuest. The highest number of journal articles were indexed as the intersection of Web of Science, Scopus, and ProQuest (11), followed by Web of Science and Scopus (10), and 'Other' with Science Direct included (9). Only two groups had mixed distribution. Scopus alone had 18 publications indexed, predominantly conference papers (12). The intersection of Web of Science and Scopus had 11 publications indexed, mostly journal articles (10). The remaining nine groups were either fully journal article (7) or conference papers dominant (2).

3.2. Content analysis results of the initial set of publications

In this section, we present the results of the content analysis. These results are segmented using a numeric breakdown for each research question, beginning with RQ1 and concluding with RQ3. Each research question represents one subsection of the content analysis, presenting the results and research contributions for each question individually.

RQ1. Do authors define bias and discrimination in research studies on AI adoption in HRM, and if so, how?

This research question aims to provide an overview of bias and discrimination definitions regarding studies on AI in HRM. Discrimination can manifest in numerous forms, ranging from unequal treatment based on personal characteristics such as gender, age, and race (Backhouse & Cherrier, 2019), (Chen, 2023a) to more contemporary discriminatory patterns such as sexual orientation, gender identity, or
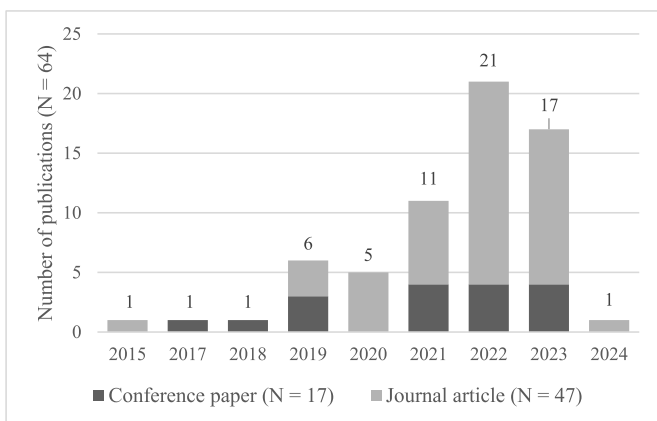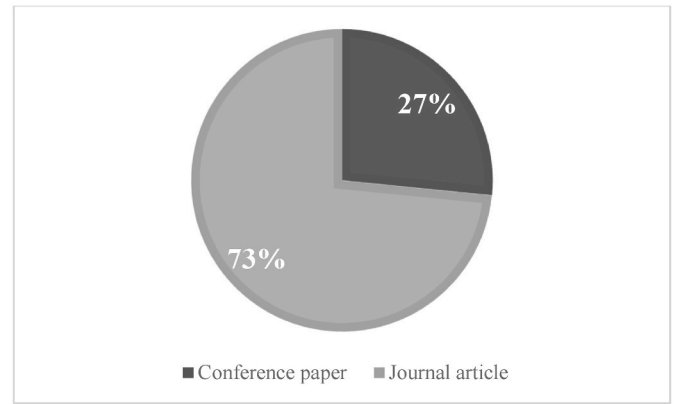

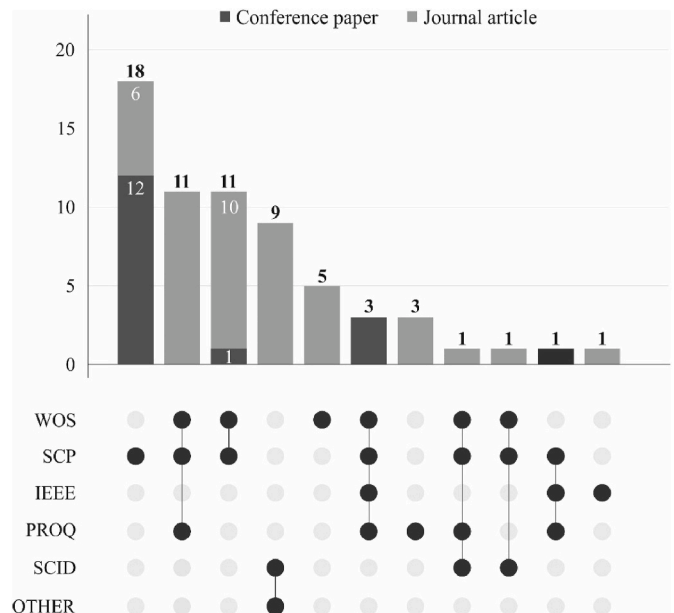Fig. 3. Distribution based on the publication type (N = 64).

Fig. 4. Publication distribution based on the database source and publication type (N = 64).

physical appearance of an individual (Badgett et al., 2021), (O' et al., 2013), (Spiegel, 2023). On the other hand, the term 'bias' is regularly used to convey technical meaning (e.g., algorithmic bias, AI bias, data bias, etc.), or it is a general term to form assumptions toward individuals or groups based on favoured preferences leading to the discriminatory actions (Fiske, 1998), (Penny), (Yam & Skorburg, 2021). Considering the diverse terminology surrounding bias and discrimination, the first step of this review is to highlight main terms discussed in the publications. For that matter, Table 7 presents types of bias and discrimination along with their respective publication sources. Furthermore, definitions of bias and discrimination are shown later in Table 9.

The most common types of discrimination were discrimination theory (Chen, 2023a), (Tambe et al., 2019), (Harris, 2020), (Kim & Bodie, 2021), (Hofeditz, Mirbabaie, et al., 2022), (Heinrichs, 2022) and indirect discrimination (Chen, 2023b), (Ramezanzadehmoghadam et al., 2021), (Aloisi, 2024), (Heinrichs, 2022), (Gaudio, 2021), (Mujtuba & Mahapatra, 2019), followed by disparate impact (Charlwood & Guenole, 2022), (Hamilton & Davison, 2022), (Mujtuba & Mahapatra, 2019), direct discrimination (Aloisi, 2024), (Heinrichs, 2022), (Gaudio, 2021) and past discrimination (Tambe et al., 2019), (Kim & Bodie, 2021), (Moore and Duffy, 2019). In most cases, all terms mentioned are defined in cited studies except indirect discrimination (Chen, 2023b),

Fig. 2. Graphical representation of included papers based on publication type and listed year.

Table 7

Representation of bias, discrimination, and related terms addressed in the analyzed publications and categorized per theme.

Theme	Term	Source
Bias	Acquiescence bias	Kimura (2023)
	Affinity bias	(Salvetti et al., 2023)*
	Age bias	(Wissemann et al., 2022)*
	Algorithmic bias	(Du et al., 2016), (Andrade, 2011), (Yam & Skorburg, 2021)*, (Kimura, 2023), (Salvetti et al., 2023), (Wissemann et al., 2022)*, (Cho et al., 2023)*, (Feldkamp et al., 2024)*, (Oravec, 2022), (Parra et al., 2022)*
	Anchoring bias	(Chen, 2023b)*
	Appraisal bias	(Shanmugam & Garg, 2015)*
	AI bias	(Varsha, 2023), (Jacob Fernandes França et al., 2023)
	Automation bias	(Shanmugam & Garg, 2015)*
	Bias (theory)	Yam and Skorburg (2021)
	Cognitive bias	(Du et al., 2016), (Yam & Skorburg, 2021)*, (Oravec, 2022)*, (Parra et al., 2022)*, (Vassilopoulou et al., 2024), (Chen, 2023b)*, (Shanmugam & Garg, 2015), (Jacob Fernandes França et al., 2023)*
	Confirmation bias	Sadler Smith et al., 2022
	Conscious bias	(Jacob Fernandes França et al., 2023)*, (Harris, 2020), (Sadler Smith et al., 2022)
	Data bias (*includes: historical data bias*, data-driven bias*, data measurement bias, data sample bias, data variable collection bias*)	(O' et al., 2013)*, (Oravec, 2022)*, (Soleimani, Intezari, Taskin, & Pauleen, 2021)*, (Veglianti et al., 2023)*, (Frissen et al., 2023)
	Defendant bias	(Varsha, 2023)*
	Designer bias	Chen (2023a)
	Gender bias	(Wissemann et al., 2022)*, (Oravec, 2022)*, (Sadler Smith et al., 2022), (Veglianti et al., 2023)*, (Gonzá et al., 2024)*, (Budhwar et al., 2023)*
	Historical bias	(Parra et al., 2022)*, (Fernández Martí et al., 2019)
	Human bias	(Oravec, 2022)*, (Robert et al., 2020)*, (Harris, 2018)*, (Kappen & Naber, 2021), (Ramezanzadehmoghadam et al., 2021)*, (Al-Alawi et al., 2021)*
	Implicit bias	(Fiske, 1998)*, (Soleimani, Intezari, Taskin, & Pauleen, 2021)*, (Gusain et al., 2023)
	Implicit racial bias	Williams (2020)
	Individual bias	(Vassilopoulou et al., 2024)*
	Intentional bias	(Shanmugam & Garg, 2015)*
	Observable bias	(Varsha, 2023)*
	Organizational bias	(Vassilopoulou et al., 2024)*
	Personal bias	(Shanmugam & Garg, 2015)*
	Popularity bias	Ramezanzadehmoghadam et al. (2021)
	Prejudice bias	Alzubaidi et al. (2023)
	Racial bias	(Wissemann et al., 2022)*, (Oravec, 2022)*, (Veglianti et al., 2023)*, (Gonzá et al., 2024)*
	Representational bias	(Ramezanzadehmoghadam et al., 2021), (Hofeditz, Mirbabaie, et al., 2022)
	Social bias (*includes: social bias, social desirability bias, societal bias*)	(Spiegel, 2023), (Kimura, 2023), (Oravec, 2022)*, (Sadler Smith et al., 2022)*, (Fernández Martí et al., 2019)
	Statistical bias	(Yam & Skorburg, 2021)*
	Structural bias	(Vassilopoulou et al., 2024)*
	System bias	(Wissemann et al., 2022)*
	Systemic bias	(Yam & Skorburg, 2021)*

Table 7 (*continued*)

Theme	Term	Source
	Unconscious bias	(O' et al., 2013)*, (Fiske, 1998)*, (Rodgers et al., 2023)*, (Tambe et al., 2019)*, (Chen, 2023b), (Jacob Fernandes França et al., 2023)*, (Sadler Smith et al., 2022), (Soleimani, Intezari, Taskin, & Pauleen, 2021)*, (Lin et al., 2021)*, (Williams, 2020)*, (Alzubaidi et al., 2023)*
	Unintentional bias	(Shanmugam & Garg, 2015)*
	Unintentional algorithmic bias	Williams (2020)
	Unfair bias	Wissemann et al. (2022)
	Unobservable bias	(Varsha, 2023)*
Discrimination	Age discrimination	(Hofeditz, Mirbabaie, et al., 2022)*
	Algorithmic discrimination	(Parra et al., 2022)*, (Hofeditz, Mirbabaie, et al., 2022)
	Discrimination (theory)	(Chen, 2023a), (Tambe et al., 2019), (Harris, 2020), (Kim & Bodie, 2021), (Hofeditz, Mirbabaie, et al., 2022), (Heinrichs, 2022)
	Disparate impact	(Charlwood & Guenole, 2022), (Hamilton & Davison, 2022), (Mujtaba & Mahapatra, 2019)
	Disparate treatment	(Kim & Bodie, 2021), (Mujtaba & Mahapatra, 2019)
	Direct discrimination	(Hofeditz, Mirbabaie, et al., 2022), (Gikopoulos, 2019)*, (Heinrichs, 2022)
	Explicit discrimination	Frissen et al. (2023)
	Gender discrimination	(Gonzá et al., 2024), (Hofeditz, Mirbabaie, et al., 2022)*
	Implicit discrimination	Frissen et al. (2023)
	Indirect discrimination	(Rodgers et al., 2023)*, (Fernández Martí et al., 2019), (Hofeditz, Mirbabaie, et al., 2022), (Gikopoulos, 2019)*, (Aloisi, 2024), (Heinrichs, 2022)
	Individual discrimination	(Heinrichs, 2022)*
	Institutional discrimination	
	Organizational discrimination	
	Past discrimination	(Tambe et al., 2019), (Kim & Bodie, 2021), (Moore and Duffy, 2019)*
	Proxy discrimination	(Hofeditz, Mirbabaie, et al., 2022), (Fernández-Martí et al., 2020)
	Racial discrimination	Williams (2020)
	„Sensitive attributes" (*age, gender, race*) discrimination	Hofeditz, Clausen, et al. (2022)
	Statistical discrimination	Chen (2023a)
	Systematic discrimination	Ramezanzadehmoghadam et al. (2021)
Other	Fairness (*algorithmic, counterfactual, distributive, individual, interactional, objective, procedural, subjective*)	(Harris, 2020), (Robert et al., 2020), (Ramezanzadehmoghadam et al., 2021), (Kim & Bodie, 2021), (Hofeditz, Mirbabaie, et al., 2022), (Danner et al., 2023), (Delecraz et al., 2022), (Mujtaba & Mahapatra, 2019)
	Justice (*informational, interpersonal, organizational, procedural*)	(Hamilton & Davison, 2022)
	Favourism; Nepotism	(Chilunjika et al., 2022), (Kshetri, 2021)*
	Broad perspective on discriminination and/or bias terminology	(Achchab et al., 2022), (Arafan et al., 2022), (Barati & Ansari, 2022), (Bartosiak & Modlinski, 2022), (Escalante et al., 2017), (Jadav et al., 2022), (Konovalova et al., 2022), (Sakka et al., 2022), (Todolí-Signes, 2019), (Vrontis et al., 2022), (Zhang & Amos, 2024)

Sources marked with '*' do not provide or discuss definitions of the mentioned terms in detail.

(Heinrichs, 2022), direct discrimination (Heinrichs, 2022) and past discrimination (Moore and Duffy, 2019). These studies discussed mentioned terms in the broad context of their respective analyses.

Other terms also included algorithmic discrimination (Vassilopoulou et al., 2024), (Aloisi, 2024), disparate treatment (Kim & Bodie, 2021), (Mujtuba & Mahapatra, 2019), gender discrimination (Gonzá et al., 2024), (Hofeditz, Mirbabaie, et al., 2022), proxy discrimination (Hofeditz, Mirbabaie, et al., 2022), (Fernández-Martí et al., 2020), explicit and implicit discrimination (Frissen et al., 2023), individual, institutional, and organizational discrimination (Heinrichs, 2022), etc. Hofeditz and colleagues (Hofeditz, Clausen, et al., 2022) introduce the concept of discrimination based on "sensitive attributes," which uniquely combines age, gender, and race under a single term. However, the term is viewed separately from gender discrimination (Gonzá et al., 2024), (Hofeditz, Mirbabaie, et al., 2022), racial discrimination (Williams, 2020), and age discrimination (Hofeditz, Mirbabaie, et al., 2022) discussed in separated studies. Among the studies on discrimination, a significant proportion (25.6%) addresses a single discrimination term, suggesting a wide range of perspectives within the field. Notably, the concepts of discrimination theory and indirect discrimination each received the highest attention in 15.4% of studies.

The most prominent terms considering bias refer to the unconscious bias (Yam & Skorburg, 2021), (Salvetti et al., 2023), (Chen, 2023b), (Shanmugam & Garg, 2015), (Sadler Smith et al., 2022), (Veglianti et al., 2023), (Gonzá et al., 2024), (Budhwar et al., 2023), (Gikopoulos, 2019), (Hemalatha et al., 2021), (Paigude et al., 2023), algorithmic bias (Soleimani, Intezari, & Pauleen, 2021), (Chen, 2023a), (Cho et al., 2023), (Feldkamp et al., 2024), (Oravec, 2022), (Parra et al., 2022), (Rodgers et al., 2023), (Tambe et al., 2019), (Varsha, 2023), (Vassilopoulou et al., 2024), cognitive bias (Soleimani, Intezari, & Pauleen, 2021), (Cho et al., 2023), (Varsha, 2023), (Vassilopoulou et al., 2024), (Harris, 2020), (Sadler Smith et al., 2022), (Soleimani, Intezari, Taskin, & Pauleen, 2021), (Veglianti et al., 2023), human bias (Varsha, 2023), (Al-Alawi et al., 2021), (Gusain et al., 2023), (Kim & Bodie, 2021), (Moore and Duffy, 2019), (Trocin et al., 2021), and gender bias (Parra et al., 2022), (Varsha, 2023), (Gonzá et al., 2024), (Fernández Martí et al., 2019), (Harris, 2018), (Kappen & Naber, 2021). These terms are discussed in 44.2% of the studies combined. Other thematic terms includes data bias (Yam & Skorburg, 2021), (Varsha, 2023), (Budhwar et al., 2023), (Fernández Martí et al., 2019), (Robert et al., 2020) which involves historical data bias, data-driven bias, data measurement bias, data sample bias, and data variable collection bias, social bias (Kimura, 2023), (Feldkamp et al., 2024), (Varsha, 2023), (Gonzá et al., 2024), (Ramezanzadehmoghadam et al., 2021), racial bias (Parra et al., 2022), (Varsha, 2023), (Fernández Martí et al., 2019), (Harris, 2018), and more. Relevant to this research, the term „artificial intelligence" (AI) bias is introduced but constitutes a smaller portion of studies (Varsha, 2023), (Jacob Fernandes França et al., 2023). Among the studies on bias, a substantial percentage (28.0%) mention a single term related to bias, indicating a broad spectrum of perspectives within the discourse on bias, mirroring the diversity observed in results on discrimination.

Besides bias and discrimination, the some studies also extensively discuss fairness, which is closely related to discrimination. In short, fairness is often viewed as the inverse of discrimination (Harris, 2020) and revolves around preventing discriminatory actions toward individuals or groups (Future of Privacy Forum, 2018) in a given social setting or situation (Getnet et al., 2014). The prominent term concerning the fairness theme is "fairness and unfairness" combined (Harris, 2020), (Ramezanzadehmoghadam et al., 2021), (Kim & Bodie, 2021), (Hofeditz, Mirbabaie, et al., 2022), (Delecraz et al., 2022), (Mujtuba & Mahapatra, 2019). Additional terms also includes algorithmic fairness (Danner et al., 2023), counterfactual fairness (Mujtuba & Mahapatra, 2019), distributive fairness (Robert et al., 2020), objective and subjective fairness (Hofeditz, Mirbabaie, et al., 2022), favoritism (Chilunjika et al., 2022), (Kshetri, 2021), justice (Hamilton & Davison, 2022), and more. Given the study's focus on bias and discrimination, we have

excluded further analysis of these terms in the following sections. A list of publications (Achchab et al., 2022), (Arafan et al., 2022), (Barati & Ansari, 2022), (Bartosiak & Modlinski, 2022), (Escalante et al., 2017), (Jadav et al., 2022), (Sakka et al., 2022), (Todolí-Signes, 2019), (Vrontis et al., 2022), (Zhang & Amos, 2024) remained 'unsorted' due to its wider perspective of the terms. Overall, out of the 164 terms analyzed, 93 (56.7%) are related to bias, 39 (23.8%) are related to discrimination, and the remaining 32 terms (19.5%) pertain to concepts such as fairness, justice, favoritism, and similar topics. Focusing exclusively on bias and discrimination, 70.5% of the identified terms address some form of bias, while 29.5% pertain to discrimination. The general representation featuring the main terms regarding bias and discrimination, along with the corresponding number of studies in which identified terms appear, is presented in Table 8 below.

The breakdown of terms has provided a foundation for addressing the primary goal of the first research question: to determine how many studies define bias, discrimination, or both within their respective contexts and how authors define both terms. While the terms bias and discrimination alone are insufficient to answer this research question, it nonetheless represents a necessary step toward reviewing definitions of bias and discrimination. To provide a clear answer to this research question, a Table 9 combines terms, definitions or explanations, and both primary and secondary sources for each definition.

Among the 58 terms related to bias and discrimination, 35 are defined or explained in enough detail (60.3%). Of the 64 studies analyzed, 30 (46.9%) included definitions or explanations for atleast one of the terms. The remaining 34 studies (53.1%) lacked clear definitions or explanations of the terms used in research. Within the discrimination theme, authors provided definitions and explanations for twelve discrimination terms, including discrimination theory (Chen, 2023a), (Tambe et al., 2019), (Harris, 2020), (Kim & Bodie, 2021), (Hofeditz, Mirbabaie, et al., 2022), (Heinrichs, 2022) and indirect discrimination (Ramezanzadehmoghadam et al., 2021), (Aloisi, 2024), (Mujtaba & Mahapatra, 2019), (Gaudio, 2021), which were among the most frequently discussed. Furthermore, studies citing algorithmic bias (Soleimani, Intezari, & Pauleen, 2021), (Chen, 2023a), (Feldkamp et al.,

Table 8
Breakdown of terms related to bias and discrimination by the amount of studies and share.

	Term	Amount	%	Total (%)
Bias	Unconscious bias	11	11.8	8.3
	Algorithmic bias	10	10.8	7.6
	Cognitive bias	8	8.6	6.1
	Gender bias	6	6.5	4.5
	Human bias	6	6.5	4.5
	Data bias	5	5.4	3.8
	Social bias	5	5.4	3.8
	Racial bias	4	4.3	3.0
	Conscious bias	3	3.2	2.3
	Implicit bias	3	3.2	2.3
	Artificial intelligence (AI) bias	2	2.2	1.5
	Historical bias	2	2.2	1.5
	Representational bias	2	2.2	1.5
	Others (n = 1)	26	28.0	19.7
	Total (B)	93	100.0	70.5
Discrimination	Discrimination (theory)	6	15.4	4.5
	Indirect discrimination	6	15.4	4.5
	Disparate impact	3	7.7	2.3
	Direct discrimination	3	7.7	2.3
	Past discrimination	3	7.7	2.3
	Algorithmic discrimination	2	5.1	1.5
	Disparate treatment	2	5.1	1.5
	Gender discrimination	2	5.1	1.5
	Proxy discrimination	2	5.1	1.5
	Others (n = 1)	10	25.6	7.6
	Total (D)	39	100.0	29.5
	TOTAL	132	-	100.0

Table 9
Definitions and explanations discussing bias and discrimination terms.

Term	Definition/Explanation	Source cited	Main source
Bias (theory)	„Within the literature on algorithmic accountability, the term bias is used in a range of different ways, often without clarity on the meaning employed. It is sometimes used to convey a specific technical meaning or it is employed as general term to mean some form of preference or unfairness."	–	Yam and Skorburg (2021)
Acquiescence bias	Occurs when respondents consistently rate all items as important, causing the distribution of responses to be skewed.	Bentler et al. (1971)	Kimura (2023)
Algorithmic bias	Occurs when an algorithmic system is trained and run on biased and unrepresentative data.	–	Chen (2023a)
	„Algorithm-based outputs of that benefit or disadvantage individuals or groups more than others without justified reason. An algorithm can be biased, for example, by training them on biased data."	Kordzadeh and Ghasemaghaei (2022)	Feldkamp et al. (2024)
	The problem of bias is associated with an AI-enhanced systems that use methods that can „unintentionally produce data that encode gender, ethnical and cultural biases."	(Bacchini & Lorusso, 2019), (Zou & Schiebinger, 2018)	Oravec (2022)
	May occur „when developers cannot formulate users' assumptions objectively or use inaccurate selection criteria."	– Tambe et al. (2019)	Tambe et al. (2019) Soleimani, Intezari, and Pauleen (2021)
	„Anomaly from machine learning algorithms caused by preconception assumptions made during the algorithm development phase or determined training data sets."	(Dilmegani)	Varsha (2023)
AI bias	Bias caused by biased training data for used by AI algorithms.	(Bolander, 2019), (Choudhury et al., 2020), (Teodorescu et al., 2020)	Jacob Fernandes Franç et al. (2023)
	„Information passed from human to AI while programming and coding the data process develops the racism and discrimination issue."	(Penny)	Varsha (2023)
Cognitive bias	„Systematic and predictable errors in judgment becoming an unintentional input into decisions."	–	Harris (2020)
	„Refer to the result of using shortcuts in thinking – heuristics."	Bazerman and Moore (2013)	Soleimani, Intezari, and Pauleen (2021)
	„Refer to a systematic error in thinking or reasoning" „Individual's deviations from rational judgement and decisions."	Leighton (2010) Ehrlinger et al. (2016)	Soleimani, Intezari, Taskin, and Pauleen (2021)
Confirmation bias	„Automatically testing a judgement by considering more confirmatory evidence than contrary evidence."	Gilovich et al. (2002)	Sadler- et al. (2022)
Conscious bias	„Perceptions of individuals or groups in society that may lead to disparate treatment."	Oates (2018)	Frissen et al. (2023)
	Perceptions influenced by factors such as „past experiences, cultural stereotypes, and social conditioning."	–	Gonzá et al. (2024)
Data bias (*data measurement bias, data sample bias, data variable collection bias*)	Caused by „how something was actually measured", data measurement bias is one source of potential bias. Data measurement is often the result of human biases. Caused by the „data sample not being representative of the broader population that the AI will interact with" is defined as data sample bias. Caused by the „decision to include or exclude specific variables in the data set."	–	Robert et al. (2020)
Designer bias	Occurs when algorithmic engineer set inappropriate goals.	(36 KE)	Chen (2023a)
Gender bias	„Gender bias refers to the unequal treatment or representation of individuals based on gender."	–	Gonzá et al. (2024)
Historical bias	„Already existing bias and sociotechnical issues."	Mozafari et al. (2020)	Ramezanzadehmoghadam et al. (2021)
Human bias	„Unfairy disadvantage toward racial minorities, women and other disadvantaged groups."	(Banks et al., 2006), (Greenwald & Krieger, 2006), (Kang & Lane, 2010), (Krieger, 1995)	Kim and Bodie (2021)
Implicit bias	„A type of automatic stereotype or prejudice that affects our opinions, decisions, and behaviors."	Brownstein (2019)	Lin et al. (2021)
Implicit racial bias	„Assumptions made about the racial identity."	Fryer and Levitt (2004)	Williams (2020)
Popularity bias	Manipulation based on popularity metrics in data.	Mozafari et al. (2020)	Ramezanzadehmoghadam et al. (2021)
Prejudice bias	„Human data bias which reflects prejudice against gender, age, race, or ideologies, leading to discriminatory predictions and recommendations."	Fernando et al. (2021)	Alzubaidi et al. (2023)
Representational bias	Input data for algorithms of certain groups or characteristics is dominant over others.	–	Hofeditz, Mirbabaie, et al. (2022)
Social bias	„Individuals or groups are disadvantaged or favoured based on their social attributes (such as age, gender, or ethnicity), which can result in discriminatory actions against them."	Fiske (1998)	Feldkamp et al. (2024)
	„Tendency to express opinions that diverge from one's true beliefs to win the approval of others or avoid their disapproval."	Crowne and Marlowe (1960)	Kimura (2023)
	Bias happens „when other people's actions coming from systems judgement."	Mozafari et al. (2020)	Ramezanzadehmoghadam et al. (2021)

(continued on next page)

Table 9 (*continued*)

Term	Definition/Explanation	Source cited	Main source
Unconscious bias	Similar to the conscious bias „but are subtle and hard to detect." *(see conscious bias above)*	–	Gonzá et al. (2024)
	„Bias that may pass unnoticed in seemingly neutral (gender) practices."	Ugarte and Rubery (2021)	Sadler Smith et al., 2022
Unintended algorithmic bias	Unintended discrimination concerns associated with the use of analytical tools that are prone to data biases.	–	Williams (2020)
Unfair bias	Bias may occur when „unrepresentative training sets are used for machine learning systems or AI-assisted decision making."	IEEE (2020)	Wissemann et al. (2022)
Adverse (disparate) impact	„Technical term referring to majority and minority group differences in the employment related opportunities, for example, hiring, promotion and termination, that are distributed as a result of using an AI system."	–	Charlwood and Guenole (2022)
	„Perpetuating historic discriminatory patterns."	–	Hamilton and Davison (2022)
	„Practices that result in a disproportionately adverse impact on protected groups."	(EEOC, 1979), (Zafar et al., 2017), (Zafar et al., 2019)	Mujtuba and Mahapatra (2019)
Disparate treatment	„Adverse decisions taken because of race, sex or any other protected class."	(42 U. S. Code et al., 2)	Kim and Bodie (2021)
	„Intentionally discriminatory practice targeted at individuals based on their protected attributes."	(EEOC, 1979), (Zafar et al., 2017), (Zafar et al., 2019)	Mujtuba and Mahapatra (2019)
Algorithmic discrimination	Occurs when using „new technologies for data processing results in harmful outcomes to the rights and freedom of natural persons."	(EU GDPR)	Aloisi (2024)
Discrimination (theory)	„Unfavorable treatment based on race, ethnicity, color, and gender that undermines employment equality."	(111 I LO, 1958), (Ruwanpura, 2008)	Chen (2023a)
	„Differential treatment based on personal characteristics, such as ethnic origin, gender, skin color, and age."	Samuelson (1952)	Chen (2023a)
	„Preference (or bias) either for or against a set of social groups that result in the unfair treatment of its members with respect to some outcome."	Bantilan (2018)	Harris (2020)
	„Distinctive form of moral wrongdoing that targets people because they belong to some group while other people which do not belong to this group are spared."	–	Heinrichs (2022)
	„Unequal treatment of different groups, not based on qualities but rather on gender, age or ethnicity."	Koolen and Van Cranenburgh (2017)	Hofeditz, Mirbabaie, et al. (2022)
	„Preferences based on age, race, color, and other protected characteristics."	EEOC (1964)	Kim and Bodie (2021)
	„Adverse actions taken based on one's demographic attributes, measured by adverse impact, evidence that any employer's decisions have a lower incidence of good and/or a higher incidence of bad outcomes than the base rate we would expect from their distribution in the relevant population."	Walsh (2015)	Tambe et al. (2019)
Direct discrimination	Occurs when „one person is treated less favorably than another is, has been, or would be treated in a comparable situation on any of the protected grounds."	(EU Law, 2000a), (EU Law, 2000b), (EU Law, 2006a), (EU Law, 2010a)	Aloisi (2024)
	„Direct discrimination occurs when a certain person is treated less favorably than another."	Craig and de Búrca (2020)	Gaudio (2021)
Gender discrimination	Consequences of unequal gender treatment.	–	Gonzá et al. (2024)
Implicit/Explicit discrimination	Occurs when „someone is treated unfairly in the same situation another person is in and preferential behavior occurs" and can occur on the basis of gender, ethnicity, sexual orientation, culture, religion, age, etc.	Österlund (2020)	Frissen et al. (2023)
Indirect discrimination	„An apparently neutral provision, criterion or practice that put persons with a membership of a protected category at a particular disadvantage compared with other persons, unless that provision, criterion or practice is justified by a legitimate aim and the means of achieving that aim are appropriate and necessary."	(EU Law, 2000c), (EU Law, 2000d), (EU Law, 2006b), (EU Law, 2010b)	Aloisi (2024)
	„Occurs when an apparently neutral provision, criterion or practice would put a person of one protected group at particular disadvantage, unless this can be objectively justified."	Craig and de Búrca (2020)	Gaudio (2021)
	„Discrimination through use of features that are implicitly defined in the dataset."	Zafar et al. (2019)	Mujtuba and Mahapatra (2019)
	Happens „when individuals appear to be treated based on seemingly neutral and non-protected attributes."	Zhang et al. (2016)	Ramezanzadehmoghadam et al. (2021)
Past discrimination	Occurs when making future predictions using or building AI tools while relying on past data.	Kim (2020)	Kim and Bodie (2021)
	Occurs when a model disproportionally leads to selection or conclusion solely based on historical data.	–	Tambe et al. (2019)
Proxy discrimination	„Discriminate in favor or against a protected class or a protected group (e.g. race, people with higher incomes and so on) but based first on legitimate or innocuous grounds."	–	Fernández-Martí et al. (2020)
	„Discrimination by proxy means that a company or an institute discriminates in favor or against minorities."	Fernández-Martí et al. (2020)	Hofeditz, Mirbabaie, et al. (2022)
Racial discrimination	Discrimination in employment based on racial characteristics of individuals or groups leading to the „fewer opportunities for career advancement, lower income, stress and poorer health."	(Pew Research Center), (Undurraga, 2019)	Williams (2020)

Table 9 (*continued*)

Term	Definition/Explanation	Source cited	Main source
„Sensitive attributes" discrimination	Discrimination based on „characteristics of candidates who are of older age, foreign race, or female and consider these attributes in the context of an AI-based system for candidate management."	–	Hofeditz, Clausen, et al. (2022)
Statistical discrimination theory	„Prejudice from assessment criteria that generalize group characteristics to individuals."	Tilcsik (2021)	Chen (2023a)
Systematic discrimination	„Patterns of behavior, policies, or practices that are part of the culture or structures of an organization, and which create or perpetuate disadvantage for racialized persons."	(Ontario Human Rights Commision)	Ramezanzadehmoghadam et al. (2021)

2024), (Oravec, 2022), (Varsha, 2023) provided most definitions of the term, followed by a social bias (Kimura, 2023), (Feldkamp et al., 2024), (Ramezanzadehmoghadam et al., 2021), and cognitive bias (Soleimani, Intezari, & Pauleen, 2021), (Harris, 2020), (Soleimani, Intezari, Taskin, & Pauleen, 2021).

In the context of this study, it is worth mentioning that authors also provided definitions for algorithmic discrimination (Aloisi, 2024), AI bias (Varsha, 2023), (Jacob Fernandes França et al., 2023), and data bias (Robert et al., 2020), which are closely related to the undesirable and unintended discriminatory actions often performed by machines against humans. AI bias, also called 'machine learning bias' or 'algorithmic bias', combines various sources of bias within the context of AI. That may include bias from algorithm predictions, training data, or cognitive biases resulting from human decision-making (IBM). Consequently, AI bias is a product of diverse factors that extend across both systemic and human aspects; therefore, we have decided to categorize AI bias as an individual class of bias.

RQ2. What are the main topics discussed in the research studies regarding bias and discrimination in HRM?

This subsection presents the findings concerning bias and discrimination topics. These topics are categorized across 12 groups, each representing a set of 'traits' based on which bias may occur, leading to unequal treatment toward individuals or groups. Two additional groups have been added for studies that approached bias and discrimination discussion more broadly and for those that didn't discuss topics in detail. For instance, bias or discrimination based on gender considers discriminatory action or preference toward an individual or group of people based on their gender identity (male, female, or other). In this context, gender is a topic within the bias and discrimination theme, and gender identity is a construct that reflects an individual's social perception or biological trait based on their sex (Blackstone et al., 2003). Table 10 summarizes publications based on main topics regarding bias and discrimination.

The most frequently discussed topic regarding bias and discrimination is gender (45.3%), followed by race (32.8%) and age (17.2%). Additionally, a significant share of publications (34.4%) address bias and discrimination from a general perspective. Although a few studies from the general discussion could potentially overlap with other publications listed under different topics, we have chosen to list them in the general discussion category because they provide a broader perspective on bias and discrimination and its occurrence in HRM. Furthermore, 8 publications (12.5%) do not delve into specific topics in detail; instead, they generally focus on other aspects of AI's potential to enhance HRM while taking care of potential discriminatory biases. Moreover, we categorized all publications according to their primary theme regarding bias and discrimination and sorted them by the topics they discuss in the publication. The following categorization distinguishes three variations: 'B' for publications focused on bias, 'D' for publications focused on discrimination, and 'BD' for publications with a balanced focus on bias and discrimination. Table 11, featuring topics and the corresponding number of studies regarding publication theme, is shown below.

Following the thematic categorization of publications, the highest number of topics is found in publications with a balanced focus on bias and discrimination (56.5%), followed by bias (36.9%) and

discrimination (6.9%). Within each theme, gender is the most prominently discussed topic (Fig. 5), comprising 24.3% of mentions in publications discussing bias and discrimination and 13.7% in total, 18.8% in publications discussing bias and 6.9% in total, and 22.2% in publications focusing solely on discrimination with only 1.5% in total. Within publications with a balanced focus on both bias and discrimination, it is noteworthy that 21.6% of topics are racial-based (12.2% in total), while 14.9% fall under general discussion (8.4% in total). General discussion also emerges as a dominant topic in both bias-focused publications with 18.8% and those focusing on discrimination 22.2%, but with only 1.5% in total. The least discussed topics across all themes are culture (B = 4.2%, D = 0%, BD = 0%), religion/faith (B = 2.1%, D = 11.1%, BD = 0%), origin (B = 2.1%, D = 0%, BD = 1.4%), and socioeconomic status (B = 2.1%, D = 0%, BD = 1.4%, each term with 3.1% mentions in total.

RQ3. *Do researchers focus more on desirable outcomes than on the undesirable effects of the use of AI?*

In this review, we identified the main terms regarding bias and discrimination. Among the 58 identified terms, 35 were defined in enough detail across 30 of the 64 studies analyzed. Furthermore, related to bias and discrimination terms, we identified twelve topics representing a set of traits that might lead to discriminatory actions, including 'others' that combine various, less popular topics. The findings present an essential element for further analysis as we conducted a set of conditions regarding the implications of AI solutions to solve discriminatory bias in HRM: (i) the study addresses bias and discrimination terms, (ii) the study provides definitions for the mentioned terms; (iii) the study discusses the positive or negative implications of AI in the context of HRM. Conditions (i) and (ii) are addressed in the first and second research questions, leaving the third condition open for analysis and discussion.

In particular, numerous studies explore the implications of AI on HRM processes, including automation in recruitment and selection, performance evaluation and prediction, talent acquisition, and overall workplace processes. Although our work refers to studies in an HRM context, the goal of this paper is not to examine operational HRM processes. This analysis is part of another paper complementary to this research, which will be published in the near future. Instead, we aim to investigate how AI may impact potential biases and discriminatory outcomes toward people within HRM.

Upon reviewing the publications on the positive and negative implications of AI adoption, we found that out of the 64 publications analyzed, 14 leaned toward positive implications (21.9%), 23 toward negative implications (35.9%), and 13 had a balanced focus on both positive and negative implications (20.3%). For the additional 14 publications (21.9%), we could not determine their main focus in discussing AI implications, as the focus was more on operational processes rather than people in HRM.

Various authors have discussed the documented history of discriminatory outcomes associated with AI solutions (Jacob Fernandes França et al., 2023), (Budhwar et al., 2023), (Moore and Duffy, 2019), (Mujtuba & Mahapatra, 2019), (Barati & Ansari, 2022). Some have even highlighted the absence of clear guidelines within organizations on how to use AI to mitigate bias and address discrimination (Gonzá et al., 2024). Furthermore, some authors have pointed out the mistaken belief that AI

Table 10

Identified topics regarding bias and discrimination.

Theme	Gender	Age	Race	Ethnicity	Culture	Disabilities	Looks/ Appearance	Sexual orientation	Religion/ Faith	Origin	Socioec. status	Other	General discussion	Not discussed in detail	Source
Bias	B		B	B	B	B									Oravec (2022)
	B	B	B						B	B					Harris (2020)
	B	B	B										B		Alzubaidi et al. (2023)
	B	B		B			B								Salvetti et al. (2023)
	B	B		B	B										Soleimani, Intezari, and Pauleen (2021)
	B			B									B		Veglianti et al. (2023)
	B		B												Harris (2018)
	B		B												Parra et al. (2022)
	B														Arafan et al. (2022)
		B													Wissemann et al. (2022)
													B		Chilunjika et al. (2022)
													B		Shanmugam and Garg (2015)
													B		Achchab et al. (2022)
													B		Al-Alawi et al. (2021)
													B		Hemalatha et al. (2021)
													B		Lin et al. (2021)
													B		Sadler Smith et al., 2022
													B		Soleimani, Intezari, Taskin, and Pauleen (2021)
													B		Trocin et al. (2021)
													B		Rodgers et al. (2023)
													B		Robert et al. (2020)
														B	Escalante et al. (2017)
														B	Gikopoulos (2019)
														B	Gusain et al. (2023)
														B	Zhang and Amos (2024)
														B	Paigude et al. (2023)
														B	Vrontis et al. (2022)
Discrimi-nation	D		D						D				D		Sakka et al. (2022)
	D					D		D							Todolí-Signes (2019)
													D		Barati and Ansari (2022)
													D		Heinrichs (2022)
Bias and discrimination	BD	BD	BD	BD		BD									Hamilton and Davison (2022)
	BD	BD	BD			BD	BD								Budhwar et al. (2023)
	BD		BD				BD	BD							Frissen et al. (2023)
	BD	BD	BD				BD								Danner et al. (2023)
	BD		BD	BD					BD						Hofeditz, Mirbabaie, et al. (2022)
	BD	BD	BD				BD	BD							Fernández-Martí et al. (2020)
	BD		BD				BD	BD							Fernández Martí et al. (2019)
	BD		BD									BD	BD		Tambe et al. (2019)
	BD	BD	BD												Hofeditz, Clausen, et al. (2022)
	BD		BD							BD					Delecraz et al. (2022)
	BD		BD			BD									Kim and Bodie (2021)
	BD		BD										BD		Chen (2023a)
	BD						BD								Kappen and Naber (2021)
	BD		BD												Ramezanzadehmoghadam et al. (2021)
	BD		BD												Varsha (2023)
	BD		BD												Vassilopoulou et al. (2024)
	BD														Feldkamp et al. (2024)

Table 10 (*continued*)

Theme	Gender	Age	Race	Ethnicity	Culture	Disabilities	Looks/Appearance	Sexual orientation	Religion/Faith	Origin	Socioec. status	Other	General discussion	Not discussed in detail	Source
	BD														Gonzá et al. (2024)
			BD												Williams (2020)
															Kshetri (2021)
												BD			Aloisi (2024)
													BD		Bartosiak and Modlinski (2022)
													BD		Charlwood and Guenole (2022)
													BD		Chen (2023b)
													BD		Jacob Fernandes França et al., 2023
													BD		Gaudio (2021)
													BD		Jadav et al. (2022)
													BD		Konovalova et al. (2022)
													BD		Moore and Duffy (2019)
													BD		Mujtaba and Mahapatra (2019)
													BD		Yam and Skorburg (2021)
														BD	Cho et al. (2023)
														BD	Kimura (2023)
Total	29	11	21	8	2	5	7	5	2	2	2	7	22	8	**Total**
%	45.3	17.2	32.8	12.5	3.1	7.8	10.9	7.8	3.1	3.1	3.1	10.9	34.4	12.5	**%**

B – bias; D – discrimination; BD – bias and discrimination.

algorithms are inherently objective and fair, which can contribute to discriminatory outcomes (Vassilopoulou et al., 2024), (Kim & Bodie, 2021). Although algorithms may 'de-humanize' bias provided by human judgment (Rodgers et al., 2023), it is indeed vulnerable to bias itself, especially if trained on inaccurate and unrepresentative data, leading to the self-reinforced biased decisions (Soleimani, Intezari, Taskin, & Pauleen, 2021). Because biased preferences can be encoded into AI solutions, decision-making may result in discriminatory outcomes toward individuals or groups of people (Yam & Skorburg, 2021).

On the other hand, research on the positive implications of AI highlights that bias is not intrinsic to the technology itself but rather influenced by human factors and the quality of the data provided. Studies suggest that well-designed AI systems have the capacity to mitigate unconscious bias, as they inherently operate with a higher degree of objectivity compared to human decision-makers (Gusain et al., 2023), (Gikopoulos, 2019). That implies that machines are not discriminatory, but rather humans who either work with biased data and unintentionally produce discriminatory outcomes or are a source of discrimination by not promoting diversity (Gonzá et al., 2024), which overflows on organizational culture and practices. Moreover, gaining knowledge about AI can lead to less bias and discrimination (Hofeditz, Clausen, et al., 2022), as AI solutions are able to provide objective decision-making, reducing the reliance on subjective heuristics (Sadler Smith et al., 2022) and removing irrelevant information that might affect decision-making (Tambe et al., 2019). By leveraging these capabilities, organizations can promote fair outcomes in domains where AI is applied.

Lastly, Table 12 combines all sources that met a set of conditions concerning the implications of AI. These conditions include discussing the terms bias and discrimination, providing definitions within their publications, and presenting both positive and negative implications of AI adoption.

The highest number of publications meeting all three conditions are categorized under negative implications (n = 11), followed by those sorted under positive implications (n = 9), and publications with balanced implications (n = 8). The most frequently defined and discussed terms include 'discrimination theory' (n = 6) and 'algorithmic bias' (n = 6). 'Discrimination theory' appears across positive (n = 1), negative (n = 3), and balanced (n = 2) implication breakdowns, while 'algorithmic bias' is predominantly found under negative implications (n = 4), followed by balanced implications (n = 2). Additionally, AI bias falls under the category of negative and balanced implications (n = 2), alongside algorithmic bias and data bias (n = 1), forming a group of negative algorithmic implications arising from inaccurate and unrepresentative (biased) data running on AI solutions.
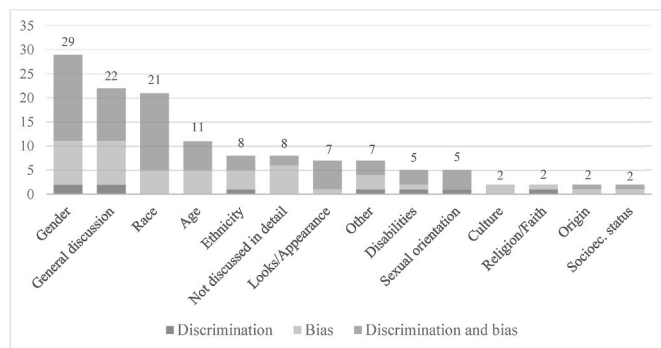
3.3. Research contributions

The first subsection (RQ1) examines the definitions of bias and discrimination in the context of using AI in HRM. From 164 identified terms, 93 (56.7%) were related to bias, 39 (23.8%) to discrimination, and 32 (19.5%) were associated with other terms such as fairness, justice, favoritism, and more. Given the study's direction described in the introduction, we have excluded further analysis of terms other than bias and discrimination. Frequency comparison between bias and discrimination revealed that 70.5% of terms are related to some form of bias, in contrast to discrimination's 29.5%. This breakdown provided a foundation for addressing the primary goal of the first research question: to determine how many studies defined or explained bias and discrimination and how authors define both terms. Of the 58 individually identified terms of bias and discrimination, 35 (60.3%) were defined (or explained) in enough detail. Among 64 studies analyzed, 30 (46.9%) included definitions of terms, whereas 34 (53.1%) lacked explicit definitions of the terms used in research. The most frequently defined terms were discrimination theory, indirect discrimination, and algorithmic bias.

Table 11
Representation of publications by theme, topics and the number of publications.

		Main theme of publications (D, B, BD)								
		Bias (B)			Discrimination (D)			Bias and discrimination (BD)		
Main topics identified in publications		Amount	% (B)	% (Total)	Amount	% (D)	% (Total)	Amount	% (BD)	% (Total)
	Gender	9	18.8	6.9	2	22.2	1.5	18	24.3	13.7
	Age	5	10.4	3.8				6	8.1	4.6
	Race	5	10.4	2.8				16	21.6	12.2
	Ethnicity	4	8.3	3.1	1	11.1	0.8	3	4.1	2.3
	Culture	2	4.2	1.5						
	Disabilities	1	2.1	0.8	1	11.1	0.8	3	4.1	2.3
	Looks/Appearance	1	2.1	0.8				6	8.1	4.6
	Sexual orientation				1	11.1	0.8	4	5.4	3.1
	Religion/Faith	1	2.1	0.8	1	11.1	0.8			
	Origin	1	2.1	0.8				1	1.4	0.8
	Socioec. status	1	2.1	0.8				1	1.4	0.8
	Other	3	6.3	2.3	1	11.1	0.8	3	4.1	2.3
	General discussion	9	18.8	6.9	2	22.2	1.5	11	14.9	8.4
	Not discussed in detail	6	12.5	4.6				2	2.7	1.5
	Total (B, D, BD)	**48**	**100.0**	**36.6**	**9**	**100.0**	**6.9**	**74**	**100.0**	**56.5**

Fig. 5. Distribution of research topics regarding bias, discrimination, or both.

The second subsection (RQ2) presents findings on bias and discrimination types categorized into 12 topics, each representing a set of traits leading to unfavorable treatment of individuals or groups. Gender was the most discussed topic (45.3%), followed by race (32.8%) and age (17.2%). More than a third of publications (34.4%) addressed bias and discrimination from a general perspective. 12.5% of publications did not delve into specific topics but discussed AI's potential to enhance HRM while taking care of discriminatory bias. Following the thematic categorization of publications, most combined bias and discrimination (56.5%), with fewer publications addressing bias alone (36.6%) or discrimination alone (6.9%). Within each theme, gender featured in the highest number, comprising 24.3% of mentions in bias and discrimination discussions (13.7% overall) and 18.8% in bias-oriented publications (6.9% overall), followed by race and age. Culture, religion/faith, origin, and socioeconomic status were the least addressed topics.

Finally, in order to explore AI's implications in addressing discrimination and mitigating bias (RQ3), the first step was to set preliminary conditions for all publications: (i) The publication addresses bias and discrimination terms; (ii) the study provides definitions for the mentioned terms; (iii) the study discusses the positive or negative implications of AI in mitigating bias and tackling discrimination. The first two conditions (i-ii) were part of RQ1 and RQ2, leaving the third for analysis. Among 64 publications, 14 leaned toward positive implications (21.9%), 23 were negative (35.9%), and 13 had a balanced focus on both positive and negative (20.3%). The remaining 14 studies (21.9%) lacked clear focus. For those studies meeting all criteria (i-iii), negative implications were most common (n = 11), followed by positive (n = 9) and balanced (n = 8). The most discussed terms were discrimination theory (n = 6) and algorithmic bias (n = 6), with discrimination theory

spanning positive, negative, and balanced views, while algorithmic bias mainly appeared in negative contexts.

4. Discussion

In this section, we highlight three topics that emerge from the research findings and that (may) warrant further discussion or research. The first relates to conceptual unclarity (and follows from the first part research), the second to conceptual diversity (and is also based upon the first part of the research), and the third to underrepresentation of types of bias and discrimination (which follows from the results of the second part of the research). Regarding the third research question, we believe that there is a good balance between research focusing on positive and negative effects, so this does not need further discussion.

4.1. Conceptual unclarity of bias and discrimination

The first point is based upon the finding that a substantial proportion of papers on bias and discrimination in the context of AI do not define (or define poorly) either term. Specifically, the systematic review found that in more than half of the papers, more precisely in 53.1%, there is no clear definition of bias and discrimination. We find this a very striking and undesirable finding, which should preferably be addressed by scholars - they could, for example, develop some standardised lexicon, or scholars could at least be more ecplicit about the concepts they use. The finding is all the more undesirable since, as indicated in the introduction, we see no reason why this review cannot be representative of other domains. In short, this result compels us to the problematic conclusion that research on bias and discriminaition in the cotnext of AI overall (and thus not only with regard to HRM) faces a major conceptual problem.

There are at least two reasons why the observed conceptual unclarity is problematic. First, the lack of conceptual clarity is not conducive to scientific inquiry. If a study does not provide a clear and precise description of bias and discrimination, it becomes (nearly) impossible, or at the very least highly challenging, to replicate the research accurately. The lack of clarity also hinders the integration of the findings with existing studies, which is crucial for advancing the field. Additionally, without well-defined terms, it becomes difficult to critically evaluate the research, as key concepts may be interpreted differently. This is particularly problematic because the processes of replication, integration, and critical evaluation are fundamental pillars of rigorous scientific research.

The second reason why the lack of clear descriptions is undesirable is that it complicates, or even makes problematic, policies around bias and discrimination. Policymakers often rely on scientific research to support

Table 12

Representation of bias and discrimination terms, publications they occur in, and thematic breakdown based on positive or negative implications of AI adoption in the context of HRM.

	Term	Source
Positive implications	Discrimination (theory)	Harris (2020)
	Disparate impact	Charlwood and Guenole (2022)
	Explicit discrimination	Frissen et al. (2023)
	Implicit discrimination	Frissen et al. (2023)
	Indirect discrimination	Ramezanzadehmoghadam et al. (2021)
	Sensitive attributes discrimination	Hofeditz, Clausen, et al. (2022)
	Systematic discrimination	Ramezanzadehmoghadam et al. (2021)
	Conscious bias	Frissen et al. (2023)
	Confirmation bias	Sadler Smith et al., 2022
	Historical bias	Ramezanzadehmoghadam et al. (2021)
	Implicit bias	Lin et al. (2021)
	Popularity bias	Ramezanzadehmoghadam et al. (2021)
	Representational bias	Ramezanzadehmoghadam et al. (2021)
	Unconscious bias	Sadler Smith et al., 2022
	Unfair bias	Wissemann et al. (2022)
Negative implications	Direct discrimination	Gaudio (2021)
	Discrimination (theory)	(Chen, 2023a), (Kim & Bodie, 2021), (Heinrichs, 2022)
	Disparate impact	Hamilton and Davison (2022)
	Disparate treatment	Kim and Bodie (2021)
	Indirect discrimination	(Gaudio, 2021), (Mujtuba & Mahapatra, 2019)
	Past discrimination	Kim and Bodie (2021)
	Artificial intelligence (AI) bias	Jacob Fernandes França et al., 2023
	Algorithmic bias	(Soleimani, Intezari, & Pauleen, 2021), (Chen, 2023a), (Feldkamp et al., 2024), (Oravec, 2022)
	Cognitive bias	(Soleimani, Intezari, & Pauleen, 2021), (Soleimani, Intezari, Taskin, & Pauleen, 2021)
	Data bias	Robert et al. (2020)
	Designer bias	Chen (2023a)
	Human bias	Kim and Bodie (2021)
	Social bias	Feldkamp et al. (2024)
	Statistical bias	Chen (2023a)
Balanced implications	Discrimination (theory)	(Tambe et al., 2019), (Hofeditz, Mirbabaie, et al., 2022)
	Gender discrimination	Gonzá et al. (2024)
	Past discrimination	Tambe et al. (2019)
	Proxy discrimination	(Hofeditz, Mirbabaie, et al., 2022), (Fernández-Martí et al., 2020)
	Acquiescence bias	Kimura (2023)
	Algorithmic bias	(Tambe et al., 2019), (Varsha, 2023)
	Artificial intelligence (AI) bias	Varsha (2023)
	Bias (theory)	Yam and Skorburg (2021)
	Conscious bias	Gonzá et al. (2024)
	Gender bias	Gonzá et al. (2024)
	Prejudice bias	Alzubaidi et al. (2023)
	Representational bias	Hofeditz, Mirbabaie, et al. (2022)
	Social bias	Kimura (2023)
	Unconscious bias	Gonzá et al. (2024)

their decisions, which is, of course, desirable. However, if studies indicate that the use of AI leads to (or can lead to) bias and discrimination, and if those studies are conceptually unclear about what bias and discrimination actually mean, it becomes difficult, if not impossible, to develop an effective policy solution. Indeed, how can one develop a solution, if the exact nature of the problem is unclear? Another possibility is that policymakers push through certain measures anyway, only to find out later that they are ineffective. This could be because they mistakenly relied on a vague definition, resulting in policies that address

a different issue and thus fail to achieve their intended goal. In the best-case scenario, such policies are merely ineffective, but there is also the risk that this could lead to decreased trust in AI, financial costs, or even that the ineffective policy inadvertently causes significant collateral damage.

4.2. Conceptual diversity of bias and discrimination

The second point for further discussion stems from the observation that there is significant conceptual diversity, meaning that the survey reveals that the terms 'bias' and 'discrimination' are used to refer to a wide range of concepts, which sometimes have little to no relation to each other. For instance, when we focus on 'bias', it becomes clear that the term can refer to the fact that the data used to train an AI system is not representative (as seen in 'data bias'), while 'bias' can also refer to the unequal treatment of individuals (as is the case with 'gender bias'). In the first case, bias is situated at the input level, specifically during the training phase, whereas in the second case, bias is situated at the output level. Although the second type of bias is usually a consequence of bias in the dataset (and not the other way around), they nonetheless pertain to different aspects. A related difference is that, in the first case, bias is a characteristic of the dataset with which the system is trained, while in the second case, bias is attributed to the system itself, particularly the output produced by that system – such as text, images, recommendations, decisions, etc. When we delve further into bias at the output level, additional distinctions can be detected. The systematic review shows that for some researchers, bias refers to a non-neutral treatment of data or information, while other scholars use 'bias' to refer to non-neutral treatment of individuals or groups. Furthermore, it appears that some researchers understand bias as a morally neutral concept, whereas in other instances, 'bias' expresses a normative concept. For example, 'human bias' is understood as 'unfair disadvantage', while 'implicit racial bias' refers merely to assumptions about racial identity. The latter is purely descriptive and does not carry any evaluative judgment about holding those assumptions.

As shown in the second part of the research, this also applies to discrimination. In this case, too, it turns out that the term is defined in numerous different ways. This is relevant information that adds value to the findings of the first part of the research. There, it was revealed that a significant portion of studies either do not define or insufficiently define 'bias' and 'discrimination', leading to the problem that comparing studies becomes very difficult, if not impossible. Now, if it turns out that studies that do provide clear definitions for both terms still do not overlap in their definitions, then these studies cannot be compared, and thus cannot be used to replicate or critique each other. This inability no longer stems from a lack of definition (as in the first part of the research) but from the observation that different studies each address something (entirely) different, even though they all deal with bias and discrimination in the context of AI. For example, when two studies suggest that AI is associated with gender bias, one cannot conclude from that alone that the studies support each other. In one case, the gender bias might be related to the fact that the dataset used to train the technology is not representative, while in the other case, the gender bias might refer to the observation that the chatbot's texts more frequently mention men than women.

4.3. Underrepresentation of bias and discrimination types

The third topic concerns the distribution of types of bias and discrimination in research on AI. The results from the second research question clearly show that this distribution is uneven, namely, the vast majority of studies on bias and discrimination in the context of AI focus on gender or race/color. More precisely, 78.1% of the studies address these two forms. This means that less than a quarter of scientific studies focus on other forms of bias and discrimination. These other types are clearly underrepresented, leading to the conclusion that, from a certain

point of view, research on bias and discrimination could itself be considered as biased.

Although we believe that neutrality in scientific research is not necessarily undesirable and is (sometimes) even impossible, it is at least a reason for concern that certain forms of bias and discrimination are so strongly represented. Indeed, it is possible that other forms of bias and discrimination have not been as thoroughly investigated. Of course, we know that an unequal number of studies does not necessarily imply that these lesser-studied forms are less well-examined than the forms with more research. The reason is that different studies often confirm the same findings, which means that the phenomenon in all its diversity is not necessarily studied more deeply than phenomena with fewer studies. Nonetheless, it remains possible that this is the case. This is a possibility that should be further explored, because if it indeed turns out that the smaller number of studies also means that these other forms of bias and discrimination are less well-understood, then there is reason to consider that undesirable.

When it becomes evident that the focus is primarily on race and gender, and consequently (much) less on other forms of bias and discrimination, such as those related to age and disabilities, the scientific knowledge of these other forms is limited. This results in an incomplete understanding of how AI systems produce undesirable effects in various contexts. From a purely scientific standpoint, this is not ideal, but this realization also has extratheoretical consequences. If bias or discrimination by AI is indeed a problem, then it is clearly an issue that needs to be addressed (or for which a solution should be sought). Such an approach benefits from a thorough understanding of the problem to be solved. However, if the knowledge is limited, the solution may also be inadequate. In short, there is a risk that strategies for addressing certain forms of bias and discrimination may not be sufficiently effective, due to an insufficient scientific foundation. If there is indeed a lack of understanding (which, as mentioned, should be further investigated), then this is a reason for the research community to shift focus and better distribute resources. This would not only increase the likelihood of effective solutions, but also eliminate the undesirable (and perhaps unintended) suggestion that certain forms of bias and discrimination are more important than others, that it is more important to thoroughly investigate racism and sexism than, for example, ableism, lookism, or ageism.

5. Conclusion

In this study, we analyzed the literature on bias and discrimination in the context of using AI in the HRM environment. The systematic review followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) framework. The main goal of this systematic review paper was to investigate relevant research on bias and discrimination in the context of using AI in the HRM environment. More specifically, we aimed to answer three questions to increase awareness of the importance of bias and discrimination controversies in using AI.

We found a lack of precise portrayal in a considerable share of papers on bias and discrimination terms. We also concluded that there's significant conceptual diversity within topics, making research on bias and discrimination even more challenging and complex. However, it adds to the allure and intrigue of the field, opening up numerous opportunities for future research. When it comes to the distribution of studies focusing on the positive and negative implications of AI, they are reasonably balanced, reflecting researchers' neutrality regarding whether the implementation of AI in HRM could lead to desirable or undesirable

effects on bias and discrimination.

Besides the earlier mentioned restriction regarding a narrowed area of research (HRM), this study has few limitations in its research approach. First, we did not use reliability or consistency measures (Cook & Beckman, 2006) to validate paper selection during the screening or eligibility phases, which could potentially lead to biased decisions regarding whether a paper was deemed eligible. Instead, considering the number of identified papers (after duplicate removal and filtering) was somewhat low, we followed a set of predefined criteria for each phase individually to aid the paper selection process and ensure consistency. Second, database (de)selection could also lead to missing other relevant studies indexed in different electronic databases. Our decision to focus on five selected databases (Web of Science, Scopus, ScienceDirect, ProQuest, and IEEE Xplore) was largely subjective and driven by the open access provided by the university. Third, qualitative text-based content analysis often contains a degree of uncertainty in determining bias and discrimination themes and topics. By addressing the limitations of this study and building upon the discoveries presented, we believe this research topic holds significant potential for generating new knowledge and closing research gaps in this emerging field.

Future research could extend our work by focusing on the types of AI addressed in studies related to bias and discrimination concerns, or by focusing more specifically on the HRM domain, where discriminatory bias is evident in areas such as hiring, promotion, training, wages, and more (Storm et al., 2023). Given the variety of AI systems available and discussed in papers, the question may be which type of AI the research addresses in the studies on bias and discrimination concerns. Some AI systems, like facial recognition or automated decision-making, may have a more deep impact on individuals and society than other technologies, including the HRM domain. Furthermore, others, ranging from AI for facial recognition to chatbots to AI that evaluates job applications, are at greater risk of being biased or discriminatory than other AI applications. Alternatively, this could involve exploring various areas within HRM, such as recruitment, selection, performance management, monitoring, or compensation strategies, to identify and address potential biases. Leading through these issues may create valuable research opportunities in the long term.

CRediT authorship contribution statement

Ivan Kekez: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Lode Lauwaert:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Nina Begičević Ređep:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Conflict of interest

There is no conflict of interest.

Appendix A

Full list of analyzed papers in alphabetical order

Authors	Publication Year	Publication Title	Paper Type	Database (*-additionally added publications)
Achchab S. & Temsamani Y.K.	2022	Use of Artificial Intelligence in Human Resource Management: "Application of Machine Learning Algorithms to an Intelligent Recruitment System"	Conference paper	SCP
Al-Alawi et al.	2021	The Role of Artificial Intelligence in Recruitment Process Decision-Making	Conference paper	WOS, SCP, IEEE, PROQ
Aloisi, A.	2024	Regulating Algorithmic Management at Work in the European Union: Data Protection, Non-discrimination and Collective Rights	Journal article	WOS
Alzubaidi, L. et al.	2023	Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements	Journal article	WOS, SCP, PROQ
Arafan A.M. et al.	2022	End-to-End Bias Mitigation in Candidate Recommender Systems with Fairness Gates	Conference paper	SCP
Barati, M. & Ansari, B.	2022	Effects of algorithmic control on power asymmetry and inequality within organizations	Journal article	WOS, SCP
Bartosiak, M.L. & Modlinski, A.	2022	Fired by an algorithm? Exploration of conformism with biased intelligent decision support systems in the context of workplace discipline	Journal article	WOS
Budhwar, P. et al.	2023	Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT	Journal article	WOS, SCP
Charlwood, A. & Guenole, N.	2022	Can HR adapt to the paradoxes of artificial intelligence?	Journal article	WOS, SCP, PROQ
Chen, Z.*	2023	Ethics and discrimination in artificial intelligence enabled recruitment practices	*Journal article*	*Nature**
Chen, Z.*	2023	Collaboration among recruiters and artificial intelligence: removing human prejudices in employment	*Journal article*	*Springer Link**
Chilunjika, A., Intauno, K. & Chilunjika, S.R.	2022	Artificial intelligence and public sector human resource management in South Africa: Opportunities, challenges and prospects	Journal article	WOS, SCP, PROQ
Cho, W., Choi, S., & Choi, H.	2023	Human Resources Analytics for Public Personnel Management: Concepts, Cases, and Caveats	Journal article	WOS, SCP, PROQ
Danner, M. et al.	2023	Towards Equitable AI in HR: Designing a Fair, Reliable, and Transparent Human Resource Management Application	Conference paper	SCP
Delecraz, S. et al.	2022	Responsible Artificial Intelligence in Human Resources Technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes	Journal article	SCP
Escalante, H.J. et al.	2017	Design of an Explainable Machine Learning Challenge for Video Interviews	Conference paper	WOS, SCP, IEEE, PROQ
Feldkamp, T. et al.	2024	Justice, trust, and moral judgements when personnel selection is supported by algorithms	Journal article	WOS, SCP
Fernández-Martínez, C. & Fernández, A.	2020	AI and recruiting software: Ethical and legal implications	Journal article	SCP
Fernández-Martínez, C. & Fernández, A.	2019	Ontologies and AI in recruiting. A rule-based approach to address ethical and legal auditing	Conference paper	SCP
França, T.J.F. et al.	2023	Artificial intelligence applied to potential assessment and talent identification in an organisational context	Journal article	WOS, SCP, SCID
Frissen, R., Adebayo, K.J. & Nanda, R.	2023	A machine learning approach to recognize bias and discrimination in job advertisements	Journal article	WOS, SCP
Gaudio, G.	2021	Algorithmic bosses can't lie! How to foster transparency and limit abuses of the new algorithmic managers	Journal article	PROQ
Gikopoulos, J.	2019	Alongside, not against: balancing man with machine in the HR function	Journal article	PROQ
Gonzales, R.*	2024	A design perspective on how to tackle gender biases when developing AI-driven systems	*Journal article*	*Springer Link**
Gusain, A. et al.	2023	E-Recruitment using Artificial Intelligence as Preventive Measures	Conference paper	SCP
Hamilton, R.H. & Davison, H.K.	2022	Legal and Ethical Challenges for HR in Machine Learning	Journal Article	WOS, SCP, PROQ
Harris, C.G.	2018	Making better job hiring decisions using "Human in the loop" techniques	Conference paper	SCP
Harris, C.G.	2020	Mitigating Cognitive Biases in Machine Learning Algorithms for Decision Making	Journal article	WOS, SCP
Heinrichs, B.*	2022	Discrimination in the age of artificial intelligence	*Journal article*	*Springer Link**
Hemalatha, A. et al.	2021	Impact of Artificial Intelligence on Recruitment and Selection of Information Technology Companies	Conference paper	SCP, IEEE, PROQ
Hofeditz, L. et al.	2022	Ethics Guidelines for Using AI-based Algorithms in Recruiting: Learnings from a Systematic Literature Review	Conference paper	SCP
Hofeditz, L. et al.	2022	Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring	Journal article	WOS, SCP, PROQ
Jadav, D. et al.	2022	EmReSys: AI-based Efficient Employee Ranking and Recommender System for Organizations	Conference paper	SCP
Kappen, M. & Naber, M.	2021	Objective and bias-free measures of candidate motivation during job applications	Journal article	WOS, SCP, PROQ
Kimura, T.	2023	Assessment of Personal Values for Data-Driven Human Resource Management	Journal article	SCP
Kim, P. & Bodie, M.*	2021	Artificial intelligence and the Challenges of Workplace Discrimination and Privacy	*Journal article*	*SSRN**

(*continued*)

Authors	Publication Year	Publication Title	Paper Type	Database (*-additionally added publications*)
Konovalova, V. et al.	2022	The impact of artificial intelligence on human resources management strategy: opportunities for the humanisation and risks	Journal article	WOS, SCP
Kshetri, N.	2021	Evolving uses of artificial intelligence in human resource management in emerging economies in the global South: some preliminary evidence	Journal article	WOS, SCP
Lin, Y. T., Hung, T.W. & Huang, L. T.L.	2021	Engineering Equity: How AI Can Help Reduce the Harm of Implicit Bias	Journal article	SCP
Moore, P.V.	2019	OSH and the future of work: Benefits and risks of artificial intelligence tools in workplaces	Conference paper	SCP
Mujtaba, D.F. & Mahapatra, N.R.	2019	Ethical Considerations in AI-Based Recruitment	Conference paper	WOS, SCP, IEEE, PROQ
Oravec, J.A.	2022	The emergence of truth machines?: Artificial intelligence approaches to lie detection	Journal article	WOS, SCP, PROQ
Paigude, S. et al.	2023	Potential of Artificial Intelligence in Boosting Employee Retention in the Human Resource Industry	Journal article	SCP
Parra, C. M., Gupta M. & Dennehy, D.	2022	Likelihood of Questioning AI-Based Recommendations Due to Perceived Racial/ Gender Bias	Journal Article	IEEE
Ramezanzadehmoghadam, M. et al.	2021	Inherent Discriminability of BERT Towards Racial Minority Associated Data	Conference paper	WOS, SCP
Robert et al.	2020	Designing fair AI for managing employees in organizations: a review, critique, and design agenda	*Journal article*	*Taylor & Francis**
Rodgers et al.*	2023	An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes	*Journal article*	*SCID**
Sadler-Smith, E., Akstinaite, V. & Akinci, C.	2022	Identifying the linguistic markers of intuition in human resource (HR) practice	Journal article	WOS, SCP, PROQ
Sakka, F. et al.	2022	Human resource management in the era of artificial intelligence: future hr work practices, anticipated skill set, financial and legal implications	Journal article	PROQ
Salvetti, F., Bertagni B. & Contardo, I.	2023	Intelligent Digital Humans for Bias-Free Recruitment Interviews: A Diversity & Inclusion Training Program	Conference paper	SCP
Shanmugam S. & Garg, L.	2015	Model employee appraisal system with artificial intelligence capabilities	Journal article	SCP
Soleimani M. et al.	2021	Cognitive biases in developing biased artificial intelligence recruitment system	Conference paper	SCP
Soleimani, M., Intezari, A. & Pauleen, D.J.	2021	Mitigating Cognitive Biases in Developing AI-Assisted Recruitment Systems: A Knowledge-Sharing Approach	Journal article	WOS, SCP, PROQ
Tambe, P., Cappelli, P. & Yakubovich, V.	2019	Artificial Intelligence in Human Resources Management: Challenges and a Path Forward	Journal article	WOS
Todoli-Signes, A.	2019	Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: the necessary collective governance of data protection	Journal article	WOS, SCP
Trocin, C. et al.	2021	How Artificial Intelligence affords digital innovation: A cross-case analysis of Scandinavian companies	Journal article	WOS, SCP, PROQ, SCID
Varsha, P.S.*	2023	How can we manage biases in artificial intelligence systems –A systematic literature review	*Journal article*	*SCID**
Vassilopoulou, J. et al.	2024	Scientism as illusio in HR algorithms: Towards a framework for algorithmic hygiene for bias proofing	Journal article	WOS
Veglianti E. et al.	2023	Customized Artificial Intelligence for Talent Recruiting: A Bias-Free Tool?	Conference paper	SCP
Williams, S.D.	2020	A textual analysis of racial considerations in human resource analytics vendors' marketing	Journal article	WOS
Wissemann, A. et al.	2022	Strategic Guidance and Technological Solutions for Human Resources Management to Sustain an Aging Workforce: Review of International Standards, Research, and Use Cases	Journal article	WOS, SCP, PROQ
Vrontis et al.*	2022	Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review	*Journal article*	*Taylor & Francis**
Yam, J. & Skorburg, J.A.	2021	From human resources to human rights: Impact assessments for hiring algorithms	Journal article	WOS, SCP
Zhang, L.X. & Amos, C.	2024	Dignity and use of algorithm in performance evaluation	Journal article	WOS, SCP

*added publications from external sources.

Data availability

Data will be made available on request.

References

111 ILO. (1958). C111 - discrimination (employment and occupation) convention. 1958 (No.111). *International Labour Organization* [Online]. Available: https://normlex.ilo. org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:C111. (Accessed 15 April 2024).

42 U.S. Code § 2000e–2. Unlawful employment practices [Online]. Available: https://www.law.cornell.edu/uscode/text/42/2000e-2. (Accessed 15 April 2024).

Achchab, S., & Temsamani, Y. K. (2022). Use of artificial intelligence in human resource management: 'Application of machine learning algorithms to an intelligent recruitment system. In L. Troiano, A. Vaccaro, R. Tagliaferri, N. Kesswani, I. Díaz Rodriguez, I. Brigui, & D. Parente (Eds.), *Lecture notes in networks and systems: Vol. 249. Advances in deep learning, artificial intelligence and robotics* (pp. 203–215). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-85365-5_20, 249.

Al-Alawi, A. I., Naureen, M., AlAlawi, E. I., & Naser Al-Hadad, A. A. (2021). The role of artificial intelligence in recruitment process decision-making. In *2021 international conference on decision aid sciences and application (DASA)* (pp. 197–203). Sakheer, Bahrain: IEEE. https://doi.org/10.1109/DASA53625.2021.9682320.

Aloisi, A. (2024). Regulating algorithmic management at work in the European union: Data protection, non-discrimination and collective rights. *International Journal of Comparative Labour Law and Industrial Relations, 40*(1), 37–70.

Alzubaidi, L., et al. (2023). Towards risk-free trustworthy artificial intelligence: Significance and requirements. *International Journal of Intelligent Systems, 2023*, 1–41. https://doi.org/10.1155/2023/4459198

Andrade, C. (2011). How to write a good abstract for a scientific paper or conference presentation. *Indian Journal of Psychiatry, 53*(2), 172. https://doi.org/10.4103/0019-5545.82558

Arafat, A. M., Graus, D., Santos, F. P., & Beauxis-Aussalet, E. (2022). End-to-End bias mitigation in candidate recommender systems with fairness gates. In *RecSys in HR'22: The 2nd workshop on recommender systems for human resources, in conjunction with the 16th ACM conference on recommender systems.* Seattle, WA, USA.

Bacchini, F., & Lorusso, L. (2019). Race, again: How face recognition technology reinforces racial discrimination. *JICES, 17*(3), 321–335. https://doi.org/10.1108/JICES-05-2018-0050

Backhouse, R. E., & Cherrier, B. (2019). Paul Samuelson, gender bias and discrimination. *The European Journal of the History of Economic Thought, 26*(5), 1053–1080. https://doi.org/10.1080/09672567.2019.1632366

Badgett, M. V. L., Carpenter, C. S., & Sansone, D. (2021). LGBTQ Economics. *The Journal of Economic Perspectives, 35*(2), 141–170. https://doi.org/10.1257/jep.35.2.141

Banks, R. R., Eberhardt, J. L., & Ross, L. (2006). Discrimination and implicit bias in a racially unequal society. *California Law Review, 94*(4), 1169. https://doi.org/10.2307/20439061

Bantilan, N. (2018). Themis-ml: A fairness-aware machine learning interface for end-to-end discrimination discovery and mitigation. *Journal of Technology in Human Services, 36*(1), 15–30. https://doi.org/10.1080/15228835.2017.1416512

Barati, M., & Ansari, B. (2022). Effects of algorithmic control on power asymmetry and inequality within organizations. *J Manag Control, 33*(4), 525–544. https://doi.org/10.1007/s00187-022-00347-6

Barredo Arrieta, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bartosiak, M. L., & Modlinski, A. (2022). Fired by an algorithm? Exploration of conformism with biased intelligent decision support systems in the context of workplace discipline. *CDI, 27*(6/7), 601–615. https://doi.org/10.1108/CDI-06-2022-0170

Bazerman, M. H., & Moore, D. A. (2013). *Judgement in managerial decision making.* Wiley.

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*(3), 186–204. https://doi.org/10.1037/h0031474

Blackstone, A. R. (2003). Gender roles and society. In J. R. Miller, R. M. Lerner, & L. B. Schiamberg (Eds.), *Human ecology: An encyclopedia of children, families, communities, and environments* (pp. 335–338). ABC-CLIO.

Bolander, T. (2019). What do we loose when machines take the decisions? *Journal of Management & Governance, 23*(4), 849–867. https://doi.org/10.1007/s10997-019-09493-x

Brownstein, M. (2019). *"Implicit bias," Stanford encyclopedia of philosophy. The metaphysics research lab, philosophy department.* Stanford, CA: Stanford University [Online]. Available: https://plato.stanford.edu/entries/implicit-bias/.

Budhwar, P., et al. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Res Mgmt Journal, 33*(3), 606–659. https://doi.org/10.1111/1748-8583.12524

Charlwood, A., & Guenole, N. (2022). Can HR adapt to the paradoxes of artificial intelligence? *Human Res Mgmt Journal, 32*(4), 729–742. https://doi.org/10.1111/1748-8583.12433

Chen, Z. (2023a). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanit Soc Sci Commun, 10*(1), 567. https://doi.org/10.1057/s41599-023-02079-x

Chen, Z. (2023b). Collaboration among recruiters and artificial intelligence: Removing human prejudices in employment. *Cogn Tech Work, 25*(1), 135–149. https://doi.org/10.1007/s10111-022-00716-0

Chilunjika, A., Intauno, K., & Chilunjika, S. R. (2022). Artificial intelligence and public sector human resource management in South Africa: Opportunities, challenges and prospects. *SA Journal of Human Resource Management, 20*. https://doi.org/10.4102/sajhrm.v20i0.1972

Cho, W., Choi, S., & Choi, H. (2023). Human resources analytics for public personnel management: Concepts, cases, and caveats. *Administrative Sciences, 13*(2), 41. https://doi.org/10.3390/admsci13020041

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal, 41*(8), 1381–1411. https://doi.org/10.1002/smj.3152

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine, 119*(2), 166.e7. https://doi.org/10.1016/j.amjmed.2005.10.036, 166.e16.

Craig, P., & de Búrca, G. (2020). *EU law: Text, cases, and materials* (7th ed.).

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. https://doi.org/10.1037/h0047358

Danner, M., Hadžić, B., Weber, T., Zhu, X., & Rätsch, M. (2023). Towards equitable AI in HR: Designing a fair, reliable, and transparent human resource management application. In D. Conte, A. Fred, O. Gusikhin, & C. Sansone (Eds.), *Communications in computer and information science: Vol. 1875. Deep learning theory and applications* (pp. 308–325). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39059-3_21, 1875.

J. Dastin, "Insight - Amazon scraps secret AI recruiting tool that showed bias against women," Reuters Accessed: July. 24, 2024. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/.

Delecraz, S., Eltarr, L., Becuwe, M., Bouxin, H., Boutin, N., & Oullier, O. (2022). Responsible Artificial Intelligence in Human Resources Technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes. *Journal of Responsible Technology, 11*, Article 100041. https://doi.org/10.1016/j.jrt.2022.100041

C. Dilmegani, "Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2022," AIMultiple. Accessed: April. 15, 2024. [Online]. Available: https://research.aimultiple.com/ai-bias/.

Du, X., Cai, Y., Wang, S., & Zhang, L. (2016). Overview of deep learning. In *2016 31st youth academic annual Conference of Chinese Association of automation (YAC), wuhan, hubei province, China* (pp. 159–164). IEEE. https://doi.org/10.1109/YAC.2016.7804882.

Dwivedi, Y. K., et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

EEOC. (1964). *Title VII of the civil rights act of 1964.* U.S. Equal Employment Opportunity Commission [Online]. Available: https://www.eeoc.gov/statutes/title-vii-civil-rights-act-1964. (Accessed 15 April 2024).

EEOC. (1979). *Questions and answers to clarify and provide a common interpretation of the uniform guidelines on employee selection procedures.* EEOC-gov [Online]. Available: https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines. (Accessed 15 April 2024).

Ehrlinger, J., Readinger, W. O., & Kim, B. (2016). Decision-making and cognitive biases. In *Encyclopedia of mental health* (pp. 5–12). Elsevier. https://doi.org/10.1016/B978-0-12-397045-9.00206-8.

Escalante, H. J., et al. (2017). Design of an explainable machine learning challenge for video interviews. In *2017 international joint conference on neural networks (IJCNN)* (pp. 3688–3695). Anchorage, AK, USA: IEEE. https://doi.org/10.1109/IJCNN.2017.7966320.

EU GDPR. Article 35(1) of Data protection impact assessment [Online]. Available: https://www.privacy-regulation.eu/en/article-35-data-protection-impact-assessment-GDPR.htm. (Accessed 15 April 2024).

EU Law. (2000a). Article 2(2)(a) of Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [Online]. Available: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32000L0043. (Accessed 15 April 2002).

EU Law. (2000b). Article 2(2)(a) of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0078. (Accessed 15 April 2024).

EU Law. (2000c). Article 2(2)(b) of Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [Online]. Available: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32000L0043. (Accessed 15 April 2024).

EU Law. (2000d). Article 2(2)(b) of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0078. (Accessed 15 April 2024).

EU Law. (2006a). Article 2(1)(a) of Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation [Online]. Available: https://eur-lex.europa.eu/eli/dir/2006/54/oj. (Accessed 15 April 2024).

EU Law. (2006b). Article 2(1)(b) of Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation [Online]. Available: https://eur-lex.europa.eu/eli/dir/2006/54/oj. (Accessed 15 April 2024).

EU Law. (2010a). Article 3(a) of Directive 2010/41/EU of the European Parliament and of the Council of 7 July 2010 on the application of the principle of equal treatment between men and women engaged in an activity in a self-employed capacity and repealing Council Directive 86/613/EEC [Online]. Available: http://data.europa.eu/eli/dir/2010/41/oj. (Accessed 15 April 2024).

EU Law. (2010b). Article 3(b) of Directive 2010/41/EU of the European Parliament and of the Council of 7 July 2010 on the application of the principle of equal treatment between men and women engaged in an activity in a self-employed capacity and repealing Council Directive 86/613/EEC [Online]. Available: http://data.europa.eu/eli/dir/2010/41/oj. (Accessed 15 April 2024).

Feldkamp, T., Langer, M., Wies, L., & König, C. J. (2024). Justice, trust, and moral judgements when personnel selection is supported by algorithms. *European Journal of Work & Organizational Psychology, 33*(2), 130–145. https://doi.org/10.1080/1359432X.2023.2169140

Fernández Martínez, C., & Fernández, A. (2019). Ontologies and AI in recruiting. A rule-based approach to address ethical and legal auditing. In *Proc. Twenty-eighth int. Joint conf. On artificial intelligence (IJCAI)* (p. 11) [Online]. Available: https://ceur-ws.org/Vol-2548/paper-06.pdf.

Fernández Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications. *Paladyn. Journal of Behavioral Robotics, 11*(1), 199–216. https://doi.org/10.1515/pjbr-2020-0030

Fernando, H., Cèsar, F., David, N., & José, H. (2021). Missing the missing values: The ugly duckling of fairness in machine learning. *International Journal of Intelligent Systems, 36*(7), 3217–3258. https://doi.org/10.1002/int.22415

Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In *The handbook of social psychology* (4th ed., pp. 357–411). McGraw-Hill.

Frissen, R., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach to recognize bias and discrimination in job advertisements. *AI & Society, 38*(2), 1025–1038. https://doi.org/10.1007/s00146-022-01574-0

Fryer, R. G., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *Quarterly Journal of Economics, 119*(3), 767–805. https://doi.org/10.1162/0033553041502180

Future of Privacy Forum. (2018). The privacy expert's guide to artificial intelligence and machine learnin. *FPF, 1400 Eye Street NW, Washington DC, USA* [Online]. Available: https://fpf.org/wp-content/uploads/2018/10/FPF_Artificial-Intelligence_Digital.pdf.

Gaudio, G. (2021). Algorithmic bosses can't lie! How to foster transparency and limit abuses of the new algorithmic managers. *Comparative Labor Law & Policy Journal Forthcoming, 35*.

Getnet, B., Jebena, T., & Tsegaye, A. (2014). The effect of employees' fairness perception on their satisfaction towards the performance appraisal practices. *International Journal of Management and Commerce Innovations, 2*(1), 174–210.

Gikopoulos, J. (2019). Alongside, not against: Balancing man with machine in the HR function. *SHR, 18*(2), 56–61. https://doi.org/10.1108/SHR-12-2018-0103

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.

González, A. S., & Rampino, L. (2024). *A design perspective on how to tackle gender biases when developing AI-driven systems*. AI Ethics. https://doi.org/10.1007/s43681-023-00386-2

González Albo, B., & Bordons, M. (2011). Articles vs. proceedings papers: Do they differ in research relevance and impact? A case study in the library and information science field. *Journal of Informetrics, 5*(3), 369–381. https://doi.org/10.1016/j.joi.2011.01.011

Graziani, M., et al. (2023). A global taxonomy of interpretable AI: Unifying the terminology for the technical and social sciences. *Artificial Intelligence Review, 56*(4), 3473–3504. https://doi.org/10.1007/s10462-022-10256-8

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review, 94*(4), 945. https://doi.org/10.2307/20439056

Gusain, A., Singh, T., Pandey, S., Pachourui, V., Singh, R., & Kumar, A. (2023). E-recruitment using artificial intelligence as preventive measures. In *2023 international conference on sustainable computing and data communication systems (ICSCDS)* (pp. 516–522). Erode, India: IEEE. https://doi.org/10.1109/ICSCDS56580.2023.10105102.

Hamilton, R. H., & Davison, H. K. (2022). Legal and ethical challenges for HR in machine learning. *Employee Responsibilities and Rights Journal, 34*(1), 19–39. https://doi.org/10.1007/s10672-021-09377-z

Harris, C. G. (2018). Aking better job hiring decisions using "human in the loop. *Presented at the HumL@ ISWC*. University of Northern Colorado, Greeley, CO 80639 USA [Online]. Available: https://ceur-ws.org/Vol-2169/paper-03.pdf.

Harris, C. G. (2020). Mitigating cognitive biases in machine learning algorithms for decision making. In *Companion proceedings of the Web conference 2020* (pp. 775–781). Taipei Taiwan: ACM. https://doi.org/10.1145/3366424.3383562.

Heinrichs, B. (2022). Discrimination in the age of artificial intelligence. *AI & Society, 37*(1), 143–154. https://doi.org/10.1007/s00146-021-01192-2

Hemalatha, A., Kumari, P. B., Nawaz, N., & Gajenderan, V. (2021). Impact of artificial intelligence on recruitment and selection of information technology companies. In *2021 international conference on artificial intelligence and smart systems (ICAIS)* (pp. 60–66). Coimbatore, India: IEEE. https://doi.org/10.1109/ICAIS50930.2021.9396036.

Hofeditz, L., Clausen, S., Rieß, A., Mirbabaie, M., & Stieglitz, S. (2022). Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. *Electron Markets, 32*(4), 2207–2233. https://doi.org/10.1007/s12525-022-00600-9

Hofeditz, L., Mirbabaie, M., Luther, A., Mauth, R., & Rentemeister, I. (2022). Ethics guidelines for using AI-based algorithms in recruiting: Learnings from a systematic literature review. In *Proceedings of the 55th Hawaii international conference on system sciences* (pp. 145–154) [Online]. Available: http://hdl.handle.net/10125/79348.

IBM, "Shedding light on AI bias with real world examples," IBM.Accessed: May 7, 2024. [Online]. Available: https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples.

IEEE. (2020). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent system. *Version 2 for Public Discussion* [Online]. Available: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.

Jacob Fernandes França, T., São Mamede, H., Pereira Barroso, J. M., & Pereira Duarte Dos Santos, V. M. (2023). Artificial intelligence applied to potential assessment and talent identification in an organisational context. *Heliyon, 9*(4), Article e14694. https://doi.org/10.1016/j.heliyon.2023.e14694

Jadav, D., et al. (2022). EmReSys: AI-based efficient employee ranking and recommender system for organizations. In *2022 international Conference on computing, communication, and intelligent systems (ICCCIS), greater noida* (pp. 440–445). India: IEEE. https://doi.org/10.1109/ICCCIS56430.2022.10037656.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electron Markets, 31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jecker, N. S., & Nakazawa, E. (2022). *Bridging East-West Differences in Ethics Guidance for AI and Robotics, 3*(3), 764–777. https://doi.org/10.3390/ai3030045. AI.

Kang, J., & Lane, K. (2010). Seeing through colorblindness: Implicit bias and the law. *UCLA Law Review, 58*, 465.

Kappen, M., & Naber, M. (2021). Objective and bias-free measures of candidate motivation during job applications. *Scientific Reports, 11*(1), Article 21254. https://doi.org/10.1038/s41598-021-00659-y

Kim, P. T. (2020). Manipulating opportunity. *Virginia Law Review, 106*(4), 867–935.

Kim, P., & Bodie, M. T. (2021). Artificial intelligence and the challenges of workplace discrimination and privacy. *35 ABA Journal of Labor and Employment Law, 289*, 289–315.

Kimura, T. (2023). Assessment of personal values for data-driven human resource management. *Data Science Journal, 22*, 22. https://doi.org/10.5334/dsj-2023-022

Konovalova, V., Mitrofanova, E., Mitrofanova, A., & Gevorgyan, R. (2022). The impact of artificial intelligence on human resources management strategy: Opportunities for the humanisation and risks. *Wisdom, 2*(1), 88–96. https://doi.org/10.24234/wisdom.v2i1.763

Koolen, C., & Van Cranenburgh, A. (2017). These are not the stereotypes you are looking for: Bias and fairness in authorial gender attribution. In *Proceedings of the first ACL workshop on ethics in natural language processing* (pp. 12–22). Valencia, Spain: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1602.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Krieger, L. H. (1995). The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity. *Stanford Law Review, 47*(6), 1161. https://doi.org/10.2307/1229191

Kshetri, N. (2021). Evolving uses of artificial intelligence in human resource management in emerging economies in the global south: Some preliminary evidence. *MRR, 44*(7), 970–990. https://doi.org/10.1108/MRR-03-2020-0168

Leighton, E. (2010). Cognitive biases. In *Encyclopedia of case study research, 2455 teller road*. Thousand Oaks California 91320 United States: SAGE Publications, Inc.. https://doi.org/10.4135/9781412957397

Lin, Y.-T., Hung, T.-W., & Huang, L. T.-L. (2021). Engineering equity: How AI can help reduce the harm of implicit bias. *Philos. Technol., 34*(S1), 65–90. https://doi.org/10.1007/s13347-020-00406-7

Moore, P. V. (2019). OSH and the future of work: Benefits and risks of artificial intelligence tools in workplaces. In V. G. Duffy (Ed.), *Lecture notes in computer science: Vol. 11581. Digital human modeling and applications in health, safety, ergonomics and risk management. Human body and motion* (pp. 292–315). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-22216-1_22, 11581.

Mozafari, M., Farahbakhsh, R., & Crespi, N. (2020). *Hate speech detection and racial bias mitigation in social media based on BERT model*. https://doi.org/10.48550/ARXIV.2008.06460

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in AI-based recruitment. In *2019 IEEE international symposium on technology and society (ISTAS)* (pp. 1–7). Medford, MA, USA: IEEE. https://doi.org/10.1109/ISTAS48451.2019.8937920.

Mujtuba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in AI-based recruitment. *Presented at the IEEE international symposium on technology in society (ISTAS)*.

Murugesan, U., Subramanian, P., Srivastava, S., & Dwivedi, A. (2023). A study of artificial intelligence impacts on human resource digitalization in industry 4.0. *Decision Analytics Journal, 7*, Article 100249. https://doi.org/10.1016/j.dajour.2023.100249

O'Brien, K. S., Latner, J. D., Ebneter, D., & Hunter, J. A. (2013). Obesity discrimination: The role of physical appearance, personal ideology, and anti-fat prejudice. *International Journal of Obesity, 37*(3), 455–460. https://doi.org/10.1038/ijo.2012.52

Österlund, P. (2020). *Preventing discrimination in recruiting through unconscious biases*. Master Thesis, LAB University of Applied Science, Finland [Online]. Available: https://www.theseus.fi/handle/10024/346242. (Accessed 15 April 2024).

Oates, M. E. (2018). Optimizing recruitment into radiology: Some simple approaches to controlling bias. *Journal of the American College of Radiology, 15*(4), 684–686. https://doi.org/10.1016/j.jacr.2017.12.002

Ontario Human Rights Commision, "Racial discrimination, race and racism (fact sheet)," Ontario Human Rights Commision. Accessed: April. 15, 2023.[Online]. Available: https://www.ohrc.on.ca/en/racial-discrimination-race-and-racism-fact-sheet.

Oravec, J. A. (2022). The emergence of 'truth machines'?: Artificial intelligence approaches to lie detection. *Ethics and Information Technology, 24*(1), 6. https://doi.org/10.1007/s10676-022-09621-6

Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Paigude, S., Pangarkar, S. C., Hundekari, S., Mali, M., Wanjale, K., & Dongre, Y. (2023). Potential of artificial intelligence in boosting employee retention in the human resource industry. *IJRITCC, 11*(3s), 1–10. https://doi.org/10.17762/ijritcc.v11i3s.6149

Parra, C. M., Gupta, M., & Dennehy, D. (2022). Likelihood of questioning AI-based recommendations due to perceived racial/gender bias. *IEEE Trans. Technol. Soc., 3*(1), 41–45. https://doi.org/10.1109/TTS.2021.3120303

L. Penny, "Robots are racist and sexist. Just like the people who created them," The Guardian. Accessed: April. 15, 2024.[Online]. Available: https://www.theguardian.com/commentisfree/2017/apr/20/robots-racist-sexist-people-machines-ai-language.

Pew Reseach Center, "Data Sources for Demographic Research," Pew Research Center. Accessed: April. 15, 2024. [Online]. Available: https://www.pewresearch.org/data-sources-for-demographic-research/.

Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods, 10*(3), 330–342. https://doi.org/10.1002/jrsm.1354

Ramezanzadehmoghadam, M., Chi, H., Jones, E. L., & Chi, Z. (2021). Inherent discriminability of BERT towards racial minority associated data. In O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. M. A. C. Rocha, E. Tarantino, & C. M. Torre (Eds.), *Lecture notes in computer science: Vol. 12951. Computational science and its applications – ICCSA 2021* (pp. 256–271). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-86970-0_19, 12951.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review, 33*(1), Article 100925. https://doi.org/10.1016/j.hrmr.2022.100925

M. Roser, "Artificial intelligence has advanced despite having few resources dedicated to its development – now investments have increased substantially," Our World in Data. Accessed: July. 24, 2024.[Online]. Available: https://ourworldindata.org/ai-investments.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Ruwanpura, K. N. (2008). Multiple identities, multiple-discrimination: A critical review. *Feminist Economics, 14*(3), 77–105. https://doi.org/10.1080/13545700802035659

Sadler Smith, E., Akstinaite, V., & Akinci, C. (2022). Identifying the linguistic markers of intuition in human resource (HR) practice. *Human Res Mgmt Journal, 32*(3), 584–602. https://doi.org/10.1111/1748-8583.12410

Sakka, F., El Maknouzi, M. E. H., & Sadok, H. (2022). Human resource management in the era of artificial intelligence: Future HR work practices, anticipated skill set, financial and legal implications. *Academy of Strategic Management Journal, 21*(1), 1–14.

Salvetti, F., Bertagni, B., & Contardo, I. (2023). Intelligent digital humans for bias-free recruitment interviews: A diversity & inclusion training program. In D. Guralnick, M. E. Auer, & A. Poce (Eds.), *Lecture notes in networks and systems: Vol. 767. Creative Approaches to technology-enhanced Learning for the Workplace and higher education* (pp. 455–461). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-41637-8_36, 767.

Samuelson, P. (1952). Spatial price equilibrium and linear programming. *The American Economic Review, 42*(3), 283–303.

Schwalbe, G., & Finzel, B. (2023). A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery*. https://doi.org/10.1007/s10618-022-00867-8

Shanmugam, S., & Garg, L. (2015). Model employee appraisal system with artificial intelligence capabilities. *Journal of Cases on Information Technology, 17*(3), 30–40. https://doi.org/10.4018/JCIT.2015070104

Soleimani, M., Intezari, A., & Pauleen, D. J. (2021). Mitigating cognitive biases in developing AI-assisted recruitment systems: A knowledge-sharing approach. *International Journal of Knowledge Management, 18*(1), 1–18. https://doi.org/10.4018/IJKM.290022

Soleimani, M., Intezari, A., Taskin, N., & Pauleen, D. (2021). Cognitive biases in developing biased Artificial Intelligence recruitment system. *Presented at the Hawaii international conference on system sciences*. https://doi.org/10.24251/HICSS.2021.620

Spiegel, T. J. (2023). Lookism as epistemic injustice. *Social Epistemology, 37*(1), 47–61. https://doi.org/10.1080/02691728.2022.2076629

Storm, K. I. L., Reiss, L. K., Guenther, E. A., Clar-Novak, M., & Muhr, S. L. (2023). Unconscious bias in the HRM literature: Towards a critical-reflexive approach. *Human Resource Management Review, 33*(3), Article 100969. https://doi.org/10.1016/j.hrmr.2023.100969

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, J. (2020). A framework for fairer machine learning in organizations. *Proceedings, 2020*(1), Article 16889. https://doi.org/10.5465/AMBPP.2020.16889abstract

Tilcsik, A. (2021). Statistical discrimination and the rationalization of stereotypes. *American Sociological Review, 86*(1), 93–122. https://doi.org/10.1177/0003122420969399

Todolí-Signes, A. (2019). Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: The necessary collective governance of data protection. *Transfer: European Review of Labour and Research, 25*(4), 465–481. https://doi.org/10.1177/1024258919876416

Triana, M. D. C., Gu, P., Chapa, O., Richard, O., & Colella, A. (2021). Sixty years of discrimination and diversity research in human resource management: A review with suggestions for future research directions. *Human Resource Management, 60*(1), 145–204. https://doi.org/10.1002/hrm.22052

Trocin, C., Hovland, I. V., Mikalef, P., & Dremel, C. (2021). How artificial intelligence affords digital innovation: A cross-case analysis of scandinavian companies. *Technological Forecasting and Social Change, 173*, Article 121081. https://doi.org/10.1016/j.techfore.2021.121081

Ugarte, S. M., & Rubery, J. (2021). Gender pay equity: Exploring the impact of formal, consistent and transparent human resource management practices and information. *Human Res Mgmt Journal, 31*(1), 242–258. https://doi.org/10.1111/1748-8583.12296

Undurraga, R. (2019). Who will get the job? Hiring practices and inequalities in the Chilean labour market. *Bulletin of Latin American Research, 38*(5), 575–590. https://doi.org/10.1111/blar.12888

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Varsha, P. S. (2023). How can we manage biases in artificial intelligence systems – a systematic literature review. *International Journal of Information Management Data Insights, 3*(1), Article 100165. https://doi.org/10.1016/j.jjimei.2023.100165

Vassilopoulou, J., Kyriakidou, O., Özbilgin, M. F., & Groutsis, D. (2024). Scientism as illusio in HR algorithms: Towards a framework for algorithmic hygiene for bias proofing. *Human Res Mgmt Journal, 34*(2), 311–325. https://doi.org/10.1111/1748-8583.12430

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education - A Multidisciplinary Journal, 23*(1), 111–127.

Veglianti, E., Trombin, M., Pinna, R., & De Marco, M. (2023). Customized artificial intelligence for talent recruiting: A bias-free tool? In C. Dal Zotto, A. Omidi, & G. Aoun (Eds.), *Lecture notes in information systems and organisation: Vol. 60. Smart technologies for organizations* (pp. 245–261). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24775-0_15, 60.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *International Journal of Human Resource Management, 33*(6), 1237–1266. https://doi.org/10.1080/09585192.2020.1871398

Walsh, D. J. (2015). *Employment law for human resource practic*. Boston, MA: Cengage Learning.

Williams, S. D. (2020). A textual analysis of racial considerations in human resource analytics vendors' marketing. *Management Research and Practice, 12*(4), 49–63.

Wissemann, A. K., Pit, S. W., Serafin, P., & Gebhardt, H. (2022). Strategic guidance and technological solutions for human resources management to sustain an aging workforce: Review of international standards, research, and use cases. *JMIR Hum Factors, 9*(3), Article e27250. https://doi.org/10.2196/27250

Xames, M. D., & Topcu, T. G. (2024). A systematic literature review of digital twin research for healthcare systems: Research trends, gaps, and realization challenges. *IEEE Access, 12*, 4099–4126. https://doi.org/10.1109/ACCESS.2023.3349379

Yam, J., & Skorburg, J. A. (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology, 23*(4), 611–623. https://doi.org/10.1007/s10676-021-09599-7

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2019). Fairness constraints: A flexible approach for fair classification. *Journal of Machine Learning Research, 20*(1), 2737–2778.

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international Conference on world wide Web, perth Australia: International world wide Web conferences steering committee* (pp. 1171–1180). https://doi.org/10.1145/3038912.3052660

Zhang, L., & Amos, C. (2024). Dignity and use of algorithm in performance evaluation. *Behaviour & Information Technology, 43*(2), 401–418. https://doi.org/10.1080/0144929X.2022.2164214

Zhang, D., et al. (2022). *The AI Index 2022 annual Report*. https://doi.org/10.48550/ARXIV.2205.03468. arXiv.

Zhang, L., Wu, Y., & Wu, X. (2016). *A causal framework for discovering and removing direct and indirect discrimination*. https://doi.org/10.48550/ARXIV.1611.07509. arXiv.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist — it's time to make it fair. *Nature, 559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Vitae: Ivan Kekez is a Ph.D. student at the Faculty of Organization and Informatics, University of Zagreb, Croatia, and a scholarship recipient under the research project 'HELA – Improving HEI maturity to implement learning analytics.' Before pursuing his Ph.D., he graduated from the Faculty of Economics, Business, and Tourism at the University of Split, Croatia.

Vitae: Lode Lauwaert is an Assistant Professor of Philosophy of Technology, and holder of the Chair Ethics and AI at KU Leuven, Belgium. His current research is on AI from the perspective of both environmental ethics and cross cultural ethics.

Vitae: Nina Begičević Ređep is a Full Professor at the University of Zagreb, Faculty of Organization and Informatics. Her areas of interest and research are business decision-making and decision theory, methods for multi-criteria decision-making, strategic planning and decision-making, educational technologies, and the application of digital technologies in education. She is the founder and holder of the Chair of the Laboratory for Strategic Planning and Decision Making at the Faculty of Organization and Informatics. She has led and participated in numerous national and international projects and co-authored over 90 research papers and books.