



OPEN ACCESS

EDITED BY

Deqing Yang,
Fudan University, China

REVIEWED BY

Aonghus Lawlor,
University College Dublin, Ireland
Shaohua Tao,
Xuchang University, China

*CORRESPONDENCE

Markus Schedl
✉ markus.schedl@jku.at

RECEIVED 23 June 2023

ACCEPTED 18 September 2023

PUBLISHED 06 October 2023

CITATION

Kumar D, Grosz T, Rekabsaz N, Greif E and Schedl M (2023) Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives. *Front. Big Data* 6:1245198. doi: 10.3389/fdata.2023.1245198

COPYRIGHT

© 2023 Kumar, Grosz, Rekabsaz, Greif and Schedl. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives

Deepak Kumar¹, Tessa Grosz², Navid Rekabsaz^{1,3}, Elisabeth Greif² and Markus Schedl^{1,3*}

¹Multimedia Mining and Search Group, Institute of Computational Perception, Johannes Kepler University Linz, Linz, Austria, ²Institute for Legal Gender Studies, Johannes Kepler University Linz, Linz, Austria, ³Human-centered AI Group, AI Lab, Linz Institute of Technology, Linz, Austria

Recommender systems (RSs) have become an integral part of the hiring process, be it via job advertisement ranking systems (job recommenders) for the potential employee or candidate ranking systems (candidate recommenders) for the employer. As seen in other domains, RSs are prone to harmful biases, unfair algorithmic behavior, and even discrimination in a legal sense. Some cases, such as salary equity in regards to gender (gender pay gap), stereotypical job perceptions along gendered lines, or biases toward other subgroups sharing specific characteristics in candidate recommenders, can have profound ethical and legal implications. In this survey, we discuss the current state of fairness research considering the fairness definitions (e.g., demographic parity and equal opportunity) used in recruitment-related RSs (RRSs). We investigate from a technical perspective the approaches to improve fairness, like synthetic data generation, adversarial training, protected subgroup distributional constraints, and *post-hoc* re-ranking. Thereafter, from a legal perspective, we contrast the fairness definitions and the effects of the aforementioned approaches with existing EU and US law requirements for employment and occupation, and second, we ascertain whether and to what extent EU and US law permits such approaches to improve fairness. We finally discuss the advances that RSs have made in terms of fairness in the recruitment domain, compare them with those made in other domains, and outline existing open challenges.

KEYWORDS

recommender system, recruitment, job recommendation, candidate recommendation, fairness, discrimination, law

1. Introduction

Recommender systems (RSs) in the recruitment domain are usable by both job seekers (job recommenders) and candidate seekers (candidate recommenders). An early application of the recruitment-related RSs (RRSs) is CASPER (Case-Based Profiling for Electronic Recruitment) (Rafter et al., 2000), an automated collaborative filtering-based personalized case retrieval system. Most modern RRSs, such as LinkedIn, have diversified their approaches and use a variety of other methods such as exploiting textual data available in the recruitment domain, or social network knowledge (Fawaz, 2019; Geyik et al., 2019). In terms of algorithms, we have come a long way from linguistics-based systems (Vega, 1990) to the current RSs which are based on deep neural networks, collaborative filtering, content-based, and knowledge-based techniques (Gutiérrez et al., 2019; Bian et al., 2020; Gugnani and Misra, 2020; Lacic et al., 2020).

However, the prevalent use of RSs has also highlighted the possibility of biased outcomes in the recruitment domain. For instance, the gender stereotypes pertaining to particular professions are observed in the current workforce (Wilson et al., 2014; Smith et al., 2021). These stereotypes can further find their way into RRSs in the form of algorithmic bias (Tang et al., 2017; Ali et al., 2019; Raghavan et al., 2020). Algorithmic bias can cause discrimination in the exposure of the job advertisement or the algorithmic hiring itself. Facebook's advertisement delivery system, for example, suffered from algorithmic gender bias while showing job advertisements (Ali et al., 2019). The Amazon hiring algorithm, infamously favoring male over female job applicants, is another real-world example (Dastin, 2018). The historical data provided to the algorithm suggested that male applicants were preferred because previously more men than women had been hired. Such behaviors prompted the industry to adopt bias mitigation in RRSs (Raghavan et al., 2020).

Such algorithmic biases can have legal consequences. Hiring decisions, whether algorithmically assisted through RRSs or solely taken by the employer, are part of the employment process and as such do not operate in a legal vacuum: Non-discrimination law plays an essential role in safeguarding from discrimination in the recruitment domain and should not be overlooked by RRS researchers. Given this context, the objective of this survey is threefold:

- to examine recent studies focusing on fairness in RRSs,
- to emphasize the significant disparities between the research conducted in the fields of computer science (CS) and law concerning this topic, and
- to identify the challenges that exist within fairness research for RSs employed in the recruitment domain.

1.1. Related surveys

There exist multiple surveys on the fairness of RSs (Zehlike et al., 2022; Deldjoo et al., 2023; Wang et al., 2023). While these surveys partly covers work on RRSs, they are not tailored to this domain, nor do they connect the discussion of fairness to the legal aspects of this domain. These two points are crucial, particularly due to the essential differences of RRSs from those used in other domains (e.g., video, music, or e-commerce): First, the decision of the system may have significant impacts on the end-users in terms of fairness and distribution of resources and can have serious legal implications. Second, the recruitment domain heavily relies on textual data of partly personal and sensitive nature, for instance, the resumes, and job posts which can contain sensitive information about candidates and employers, respectively. There also exist several surveys on RRSs, e.g., Dhameliya and Desai (2019), de Ruijt and Bhulai (2021), Freire and de Castro (2021), and Thali et al. (2023), but only de Ruijt and Bhulai (2021) include a dedicated section covering fairness aspects. However, de Ruijt and Bhulai (2021) provide a broad scope on the topic, and only present high-level insights on fairness aspects. In particular, we are not aware of any survey in this domain that draws the connection between algorithmic aspects of RRSs and the intertwined legal

TABLE 1 List of abbreviations used in the article. For abbreviations of fairness metrics, please refer to Table 3.

AI	Artificial Intelligence
CBF	Content-based Filtering
CF	Collaborative Filtering
CJEU	Court of Justice of the European Union
CS	Computer Science
DP	Demographic Parity
EO	Equal Opportunity
IF	Individual Fairness
KB	Knowledge-Based
PF	Proportional Fairness
RRS	Recruitment-related Recommender System
RS	Recommender System

facets. To fill these gaps in existing surveys on the fairness of RSs and RRSs, the survey at hand calls for attention to the research specifically addressing the fairness of RSs in the recruitment domain, with a multidisciplinary analysis from both technical and legal perspectives.

1.2. Literature search

To identify relevant literature, we conducted a series of searches on DBLP¹ with the keywords “Job”/“Candidate” in conjunction with “Fair”/“Bias” to create a candidate list of publications for this survey. Subsequently, we selected a subset of this list based on their relevance to RSs in the recruitment domain after a careful manual inspection of each paper. We further enriched the list with articles from the workshop series on Recommender Systems for Human Resources (RecSys in HR) and with further recent works spotted by studying the references in the collected papers. The surveyed literature covers the work published until May 2023.

1.3. Outline

The survey is structured as follows: Section 2 provides an overview of the current research on fairness in RRSs. In Section 3, we focus on the significant legal aspects that pertain to RSs in the recruitment domain, aiming to bridge the gap between the legal and technical dimensions of ensuring fairness in candidate recommenders, and to highlight the shortcomings of the existing approaches for fairness in RS. Finally, Section 4 discusses the open challenges and future directions related to fairness research in RRSs. For easy reference, Table 1 contains a list of abbreviations used throughout the article.

¹ <https://dblp.org/> (accessed May 2023).

TABLE 2 Overview and categorization of the research works surveyed.

Article	RS		Dataset(s)	Fairness measurement		Debiasing strategy	Sensitive attribute(s)
	Type	Algorithm		Definition(s)	Metric(s)		
Li et al. (2023)	Job	CF	Private	DP	UGF	Post	Gender (G)
Rus et al. (2022)	Job	CBF	Private	EO, DP	TPRP, SAT	Pre, In	G
Ntioudis et al. (2022)	Job	KB	Synthetic, private	—	—	Pre	Migrant
Zhang (2021)	Job	—	Synthetic	DP	SDR, LDR	—	G
Shishehchi and Banhashem (2019)	Job	KB	Private	—	UAT	Pre	Disability
Tang et al. (2017)	Job	—	Scraped	—	MWU	—	G
Scher et al. (2023)	Candidate	—	Synthetic	EO, DP	BGSD, CF, TNR	Post	—
Jourdan et al. (2023)	Candidate	CBF	BIOS	EO	TPRP	Pre, In	G
Delecraz et al. (2022)	Candidate	—	Private	EO	TPRP, DI, SP	—	Multi
Markert et al. (2022)	Candidate	CBF	Synthetic	IF	IFC	—	G, Marital status
Bei et al. (2020)	Candidate	Other	Synthetic	PF	Violations	Post	G, Region
Arafan et al. (2022)	Candidate	—	Synthetic, Private	DP	NDKL	Pre, Post	G
Tran et al. (2022)	Candidate	KB	ACI, TO, HR	—	—	Post	Disability
Syed and Shivendu (2022)	Candidate	—	—	EO, DP	TPRP	Pre	—
Burke et al. (2021)	Candidate	Other	SIOP2021	EO	AIR	Post	—
Tran et al. (2021)	Candidate	KB	Synthetic, Private	—	—	Post	Disability
Wilson et al. (2021)	Candidate	—	Private	EO, DP	AIR, MWU, KWH	—	G, Ethnicity
Elbassuoni et al. (2020)	Candidate	—	Synthetic, Private	DP	EMD	—	G, Ethnicity
Elbassuoni et al. (2019)	Candidate	—	Private	DP	EMD	—	Multi
Geyik et al. (2019)	Candidate	—	Synthetic, Private	EO, DP	NDKL, MS, II, IC	Post	G, Age
Chen et al. (2018)	Candidate	—	Scraped	DP, IF	MWU, TDRC, EC	—	G
Amer-Yahia et al. (2020)	Both	—	Synthetic, Private	DP	KT, EMD, Exposure	—	G, Ethnicity, Nationality

In RS Type, “Both” is used when both job recommender and candidate recommender are used, and for Sensitive Attribute(s), “Multi” is used when there are four or more attributes (e.g., Gender, Nationality, Age, Ethnicity, Language). The symbol “—” is used when the column is not applicable or cannot be inferred from the paper.

2. Current state of research

As a core reference for interested readers, we provide in Table 2 an overview and categorization of existing research that addresses fairness in RRSs. The individual works are classified based on the datasets utilized for the experiments, the aspired fairness definitions, the metrics employed to assess fairness, the stage in the pipeline where debiasing strategies are implemented to achieve fairness, and the sensitive attributes explored during the experiments.

This section analyzes the current state of fairness research in RRSs. Our examination of the existing literature is structured around various aspects, including the *recommendation algorithm* underlying the RRSs (Section 2.1), the *datasets* used in the reviewed works (Section 2.2), the *definition of fairness* (Section 2.3), the *metrics* used to evaluate fairness or unfairness in the datasets and created recommendations (Section 2.4), *evidence of bias*

within these algorithms (Section 2.5), and the different *approaches* explored to attain fairness (Section 2.6).

In the article at hand, we adopt the terminology common in RRSs research, i.e., we refer to “users” (U) and “items” (I), the former representing the entity for which recommendations are made (input), the latter the recommendations themselves (output). Hence, for a job recommender system, the user is the candidate looking for a job, and the items are job posts. In contrast, for a candidate recommender, the user is the job post and the items are candidates.

2.1. Recommendation algorithms in the recruitment domain

RRSs can be classified into four categories: collaborative filtering (CF), content-based filtering (CBF), knowledge-based

(KB) recommenders, and hybrid recommenders. In research on fairness of RRSs, company platforms are mostly investigated, e.g., Indeed, Monster, and CareerBuilder² in Chen et al. (2018); LinkedIn in Tang et al. (2017) and Geyik et al. (2019); TaskRabbit and Google Job Search³ in Amer-Yahia et al. (2020) and Elbassuoni et al. (2020). These works are about investigating fairness of these popular platforms rather than improving it.

In contrast to the proprietary and commonly non-disclosed recommendation algorithms used by the aforementioned services, RRS approaches found in the surveyed literature can be categorized into the following:

- Collaborative Filtering (CF): CF algorithms employed in RRSs are all based on the similarity between users, calculated through the users' interactions with the items. In the RRSs literature, users are always job seekers, i.e., no CF approaches are used for candidate recommenders, only job recommenders. The adopted algorithms include probabilistic matrix factorization (Salakhutdinov and Mnih, 2007), neural matrix factorization (He et al., 2017), session-based model STAMP (Short-Term Attention/Memory Priority) (Liu et al., 2018), and matrix factorization with global bias terms (Koren et al., 2009).
- Content-based Filtering (CBF): CBF algorithms are based on the similarity between the items and the users, most commonly implemented as direct matching of users and items through text-based similarity. In the surveyed RRSs literature, users can be job posts (candidate recommenders) or candidates (job recommenders). Rus et al. (2022) and Jourdan et al. (2023) implement RoBERTa (Liu et al., 2019) and BERT (Devlin et al., 2019) language models, respectively, to learn candidate-job similarity. Markert et al. (2022) use a custom regression model to learn the candidate-job similarity. In these works, both language model and regression model are neural network-based models.
- Knowledge-based (KB): The algorithms in this category utilize domain-specific knowledge to create ontology and compute similarity between user-item pairs. Existing works that fall in this category use directed acyclic causal graphs (Tran et al., 2021, 2022) and knowledge representation using domain ontology (Shishchchi and Banihashem, 2019; Ntioudis et al., 2022).
- Hybrid: Hybrid approaches utilize combinations of the previous approaches (Luo et al., 2019). No such hybrid could be found in the conducted literature search.
- Others: We group here the studies that do not use the mentioned recommendation approaches. In particular, Burke et al. (2021) implements spatial search-based candidate selection, and Bei et al. (2020) deploys integer linear programming-based candidate selection.

² <https://www.indeed.jobs/>, <https://www.monster.com/>, <https://www.careerbuilder.co.uk/> (accessed May 2023).

³ <https://www.taskrabbit.com/>, <https://jobs.google.com/about/> (accessed May 2023).

2.2. Datasets

In Table 2, we divide the datasets into four categories: dataset that cannot be accessed or recreated (denoted *Private* in the table), scraping details are given (*Scraped*), procedure to create the artificial dataset is given (*Synthetic*), and name of the dataset if it is public. The datasets used in the reviewed works are primarily private or synthetically created. Table 2 also reveals the surprising absence of popular public job recommender datasets, such as the datasets used in the ACM Recommender Systems Challenge 2016 and 2017 (Abel et al., 2016, 2017) and the Career Builder 2012 dataset.⁴ Public datasets are used only for candidate recommenders.

- BIOS (De-Arteaga et al., 2019): This dataset has been created by scrapping biographies using Common Crawl.⁵ The dataset contains biographies of individuals with the attributes current job and gender. The dataset proposed by the authors contains 397,340 biographies with 28 different occupations.
- Adult Census Income (ACI) (Becker and Kohavi, 1996): The ACI dataset has been extracted from the 1994 USA census database and includes 48,842 individuals' records. For each individual, it contains features such as occupation, age, education, marital status, salary, race, and gender.
- TO⁶: The dataset contains 1,129 Russian workers' income and 16 features describing them, such as gender, age, profession, experience, industry, employee turnover, supervisor, supervisor's gender, and recruitment route.
- HR⁷: This dataset contains 15,000 individuals' retention records with the features satisfaction level, last evaluation, average monthly hours, work accident, salary, and time spent in company.
- SIOP2021 (Koenig and Thompson, 2021): This dataset was introduced in the machine learning challenge of the Annual Conference of the Society for Industrial and Organizational Psychology 2021. The dataset contains the three attributes performance, turnover data, and the protected group membership⁸ for 7,890 respondents.

2.3. Fairness definitions for recruitment-related recommender systems

We review different fairness definitions in RRSs in this section. Overall, as shown in Table 2, one of the most used fairness definitions in the context of RRSs is demographic parity (DP). Also, the fairness definition for job recommenders is restricted to DP in the identified literature. Overall, DP is the

⁴ <https://www.kaggle.com/competitions/job-recommendation/overview> (accessed June 2023).

⁵ <https://commoncrawl.org/> (accessed June 2023).

⁶ <https://www.kaggle.com/datasets/davinwijaya/employee-turnover> (accessed June 2023).

⁷ <https://www.kaggle.com/datasets/liujiqi/hr-comma-sepcsv> (accessed June 2023).

⁸ The actual protected group is not mentioned by the creators.

most often used fairness definition for both job and candidate recommenders. We will see in the definition of DP that it does not consider the quality of recommendation (i.e., whether relevant items are ranked high or low is ignored. This results in an easier adaptation of this fairness definition from classification to recommendation tasks for both candidate recommenders and job recommenders.

Most fairness definitions mentioned in the literature are adapted from the binary classification setting to the RS setting. The core fairness definitions from the classification are listed below:

- *Demographic parity (DP)*: A binary predictor \hat{Y} is said to satisfy demographic parity with respect to protected attribute $A \in \{a_1, \dots, a_l\}$ that can take l values if \hat{Y} is independent of A (Dwork et al., 2012).

$$P(\hat{Y}|A = a_1) = \dots = P(\hat{Y}|A = a_l)$$

- *Equal opportunity (EO)*: A binary predictor \hat{Y} is said to satisfy equal opportunity with respect to a protected attribute $A \in \{a_1, \dots, a_l\}$ that can take l values and ground truth Y if they are independent conditional on the ground truth outcome being favorable (Hardt et al., 2016).

$$P(\hat{Y} = 1|A = a_1, Y = 1) = \dots = P(\hat{Y} = 1|A = a_l, Y = 1)$$

- *Individual fairness (IF)*: A predictor \hat{Y} is said to satisfy individual fairness if for data-point x from dataset D for all x' that are similar to x (i.e., all their attributes are the same except for the sensitive attribute) the predictor predicts the same class (Ruoss et al., 2020).

$$x \sim D, \forall x' \in \mathbb{R} : \phi(x, x') \implies \mu(\hat{Y}_x, \hat{Y}_{x'})$$

where $\phi(x, x') = 1$ iff x and x' are similar and $\mu(\hat{Y}_x, \hat{Y}_{x'}) = 1 \iff \hat{Y}_x = \hat{Y}_{x'}$

As we explain in the following, the fairness definitions for candidate recommenders are strongly aligned with classification scenarios, i.e., whether an item (candidate) occurs in the recommendation list or not. At the same time, fairness definitions for job recommenders require further modification. The fairness definitions in RRSs are always from the candidate's perspective. In the case of job recommenders, the involved variables are sensitive attributes A_u (e.g., gender or ethnicity) of the input u (i.e., candidate), and the recommended list Q_u (i.e., list of job posts). For candidate recommenders, the sensitive attribute A_i belongs to the item i in the recommended list Q_u (i.e., list of candidates). Furthermore, defining fairness for job recommenders requires the function \mathcal{F} defined over the recommended list Q_u (list of job posts), which measures the quality of recommendation (e.g., through precision or recall metrics) or some other property of Q_u like the average salary of jobs in Q_u . Based on these definitions, in the following, we review the fairness definitions for RRSs, adapted from classification tasks:

- *DP for job recommender (Li et al., 2023)*: A job recommender satisfies DP for sensitive attribute gender

$A_u \in \{male, female, non - binary\}$ if some measure, expressed as a function \mathcal{F} defined over the recommendation list of jobs, is independent of the value of A_u , i.e.,

$$P(\mathcal{F}(Q_u)|A_u = male) = P(\mathcal{F}(Q_u)|A_u = female) = P(\mathcal{F}(Q_u)|A_u = non - binary)$$

Example: The average salary of jobs recommended have the same distribution for the male, female, and non-binary candidate groups.

- *DP for candidate recommender (Geyik et al., 2019)*: A candidate recommender satisfies DP for attribute tuple $A_i = \langle gender_i, age_i \rangle$, if the existence or absence of any candidate i in the recommended list Q_u for job ad u is independent of their attributes A_i .

$$P(i \in Q_u | A_i = \langle gender_i, age_i \rangle) = P(j \in Q_u | A_j = \langle gender_j, age_j \rangle)$$

$$P(i \notin Q_u | A_i = \langle gender_i, age_i \rangle) = P(j \notin Q_u | A_j = \langle gender_j, age_j \rangle)$$

Example: A young male has the same probability as an old female of being included in the recommendation list for the job.

- *EO for candidate recommender (Geyik et al., 2019)*: A candidate recommender satisfies EO with respect to protected attribute tuple $A_i = \langle gender_i, age_i \rangle$ if candidate i 's existence in the recommended list Q_u with respect to protected attributes A_i is conditionally independent of the candidate being qualified [i.e. $\rho(i, u) = 1$] for the job u .

$$P(i \in Q_u | A_i = \langle gender_i, age_i \rangle, \rho(i, u) = 1) = P(j \in Q_u | A_j = \langle gender_j, age_j \rangle, \rho(i, u) = 1)$$

Example: A young male who satisfies the requirements for the job has the same probability of being selected for the job as an old female that satisfies the same requirements.

- *IF for candidate recommender (Markert et al., 2022)*: A candidate recommender satisfies IF if for an item i (i.e., candidate) in the recommended list Q_u (for a job ad u) all the candidates i' similar to i (i.e., their attributes are the same except for the sensitive attribute) in the recommended lists have nearby positions in the ranking.

$$i \in Q_u, \forall i' \in Q_u : \phi(i, i') \implies \mu(Q_u, i, i')$$

where $\phi(i, i') = 1$ iff i and i' are similar and $\mu(Q_u, i, i') = 1 \iff pos(i, Q_u) \approx pos(i', Q_u)$

Example: If two candidates of different gender in the item set have the same attributes, they should be ranked at nearby positions in the recommendation list.

In contrast to the mentioned definitions, *Proportional Fairness (Bei et al., 2020)* is not adapted from classification and is directly defined for RRSs.

- **Proportional Fairness (PF)** (Bei et al., 2020): For a candidate recommender to satisfy PF, the selected set of candidates S and candidate attribute set $A = \{a_1, \dots, a_l\}$, at least fraction α_j and at most fraction β_j of S have attribute a_j .

Example: Five candidates are selected for a given job by the RRS. The attributes of these people are gender (male vs. female) and region (Europe vs. Africa). Then, for proportional fairness over the attributes male, female, Europe, and Africa under the constraints (least fraction, most fraction), respectively, (0.4, 0.6), (0.4, 0.6), (0.2, 0.4), and (0.2, 1.0), it is required that the number of people out of 5 with attribute male, female, Europe, and Africa are $\{2, 3\}, \{2, 3\}, \{1, 2\}$, and $\{1, 2, 3, 4, 5\}$, respectively.

2.4. Fairness and unfairness metrics

Formalizing the fairness definitions introduced above, the RRSs literature has proposed various metrics to quantify the degree to which fairness is achieved by a given RRS. In this section, we therefore present the fairness metrics used in the surveyed literature. As we could already see in Table 2, they are very diverse. The only recurring metrics are TRPR, EMD, NDKL, MWU, and AIR. The frequency of each metric’s use in literature can be seen in Table 3, along with a categorization of the metrics with respect to the targetted fairness definitions. Below we introduce the common metrics and the metrics we refer to later in the manuscript. For the remaining ones, we refer the reader to the corresponding references provided in Table 2.

The metrics in the literature addressing DP consider the distribution of sensitive attributes over recommended lists, as shown for here:

- **User Group Fairness (UGF)** (Li et al., 2023): Given metric \mathcal{F} (e.g., average salary) over a recommended list of jobs Q_u for the user u (i.e. candidate) and the set of all candidates U partitioned in two mutually exclusive sets U_{a_1}, U_{a_2} based on protected attribute $A \in \{a_1, a_2\}$, UGF is defined as:

$$UGF(U_{a_1}, U_{a_2}, Q) = \left| \frac{1}{|U_{a_1}|} \sum_{u \in U_{a_1}} \mathcal{F}(Q_u) - \frac{1}{|U_{a_2}|} \sum_{u \in U_{a_2}} \mathcal{F}(Q_u) \right|$$

where $Q = \{Q_u \forall u \in U\}$ is the set of all recommendation lists.

- **Set Difference Rate (SDR)** (Zhang, 2021): SDR is defined to measure proportion of attribute-specific jobs. SDR between set of items (i.e., jobs) I_{a_1} and I_{a_2} recommended only to users (i.e., candidates) with sensitive attribute $A = a_1$ and $A = a_2$ (where $A \in \{a_1, a_2\}$) respectively and the set of all jobs I .

$$SDR(I_{a_1}, I_{a_2}) = \frac{|I_{a_1}| + |I_{a_2}|}{|I|}$$

- **List Difference Rate (LDR)** (Zhang, 2021): LDR is defined to measure differences in ranking due to the binary protected attribute $A \in \{a_1, a_2\}$ of a user. LDR for a pair of recommendation lists Q_u and $Q_{\hat{u}}$ of a candidate u and its

TABLE 3 Metrics categorized according to the fairness definition associated, with frequency of papers using the metric.

Fairness definition	Metric	Frequency
Demographic Parity (DP)	Earth Mover’s Distance (EMD)	3
	Mann-Whitney U test (MWU)	3
	Normalized KL Divergence(NDKL)	2
	User Group Fairness (UGF)	1
	Salary Association Test (SAT)	1
	Set Difference Rate (SDR)	1
	List Difference Rate (LDR)	1
	Min/Max Skew (MS)	1
	Kendall’s Tau (KT)	1
	Kruskal-Wallis H test (KWH)	1
Total Difference between the Recall Curves (TDRC)	1	
Equal opportunity (EO)	True Positive Rate Parity (TPRP)	4
	Adverse Impact Rate (AIR)	2
	Counterfactual Fraction (CF)	1
	True Negative Rate (TNR)	1
	Infeasible Index (II)	1
	Infeasible Count (IC)	1
	Exposure	1
Individual Fairness (IF)	Individual Fairness Certification (IFC)	1
	Effect Coefficient (EC)	1
Proportional Fairness (PF)	Violations	1
—	User Acceptance Test (UAT)	1
	Between Group Skill Difference (BGSD)	1
	Disparate Impact (DI)	1
	Statistical Parity (SP)	1

counterfactual \hat{u} (created by changing u ’s protected attribute), respectively, is

$$LDR(Q_u, Q_{\hat{u}}) = \frac{\sum_{j=1}^{|Q_u|} \gamma(Q_u, Q_{\hat{u}}, j)}{|Q_u|}$$

where $\gamma(Q_u, Q_{\hat{u}}, j) = 1$ if the j^{th} job ad in Q_u and $Q_{\hat{u}}$ is the same.

- **Earth Mover’s Distance (EMD)** (Pele and Werman, 2009): EMD can be defined for sensitive attribute $A \in \{a_1, \dots, a_l\}$ that can take l values by dividing the ranking Q_u in two groups, group G with $A = a_1$ and group \bar{G} with $A \neq a_1$. EMD between the two groups represents the smallest amount of required change in the ranking scores of the bigger group to obtain the ranking scores of the smaller group. Then, EMD between G and \bar{G} for

the smallest amount of change in the ranking score scenario is

$$EMD(G, \bar{G}) = \frac{\sum_{i,j} f_{i,j} |i - j|}{\sum_{i,j} f_{i,j}}$$

where $f_{i,j}$ is the change in i^{th} rank score of the bigger group to get the j^{th} rank score of the smaller group. To give an example, assume that the two groups are young G and old \bar{G} candidates. Groups are represented by the score their members have at each rank. So, for $G = \{0, 0.3, 0.2\}$ and $\bar{G} = \{0.5, 0, 0\}$, the solution of smallest change is to rerank G to get the same score distribution as \bar{G} . Here, $EMD(G, \bar{G}) = \frac{0.3*1+0.2*2}{0.3+0.2} = 1.4$.

- **Normalized Discounted Kullback-Leibler Divergence (NDKL)** (Geyik et al., 2019): Given ranked list Q_u and two distributions $\mathcal{P}_{A_{Q_u}}$ and \mathcal{P}_A , which are distributions of attribute A in Q_u and the desired distribution of A , respectively, the NDKL is

$$NDKL(Q_u) = \frac{1}{Z} \sum_{j=1}^{|Q_u|} \frac{1}{\log_2(j+1)} d_{KL}(\mathcal{P}_{A_{Q_u}} || \mathcal{P}_A),$$

$$Z = \sum_{j=1}^{|Q_u|} \frac{1}{\log_2(j+1)}$$

where d_{KL} is KL-divergence. It should be noted that $NDKL = 0$ for all users would imply demographic parity is achieved.

- **Mann-Whitney U (MWU) test** (Corder and Foreman, 2014): MWU test is a non-parametric statistical test where the null hypothesis for job recommender is:
 - *For candidate recommender*: The ranks of candidates with a particular protected attribute is not significantly different from the ranks of candidates with another attribute.
 - *For job recommender*: The recommended list of job posts is not significantly different for candidates with different protected attributes.

The metrics associated with EO are dependent on the quality of recommendations and the distribution of sensitive attributes over recommended lists.

- **True Positive Rate Parity (TPRP)** (Delecraz et al., 2022): TPRP in candidate recommendation for a given binary protected attribute $A \in \{a_1, a_2\}$ and recommendation list Q_u for job u is defined as

$$TPRP(u) = |P(x \in Q_u | A_x = a_1, \rho(u, x) = 1) - P(x \in Q_u | A_x = a_2, \rho(u, x) = 1)|$$

where $\rho(u, x) = 1$ implies that a candidate x sampled randomly from the candidate set is suitable for job u , and $TPRP = 0$ implies EO is achieved. TPRP is also called sourcing bias (Syed and Shivendu, 2022) and true positive rate gap (Jourdan et al., 2023) in the literature.

- **Adverse Impact Ratio (AIR)** (Burke et al., 2021): AIR for binary sensitive attribute $A \in \{a_1, a_2\}$ and recommended list Q_u for job u is,

$$AIR(u) = \frac{|P(x \in Q_u | A_x = a_1, \rho(u, x) = 1)|}{|P(x \in Q_u | A_x = a_2, \rho(u, x) = 1)|}$$

where, $\rho(u, x) = 1$ implies that a candidate x sampled randomly from the candidate set is suitable for job u , and $AIR = 1$ implies EO is achieved.

There also exist a few metrics which are not associated with a particular fairness definition but are nevertheless related with fairness, e.g., UAT measures the disadvantaged groups' perception of the RRS, and several metrics quantify the fairness of RRS datasets.

- **User Acceptance Test (UAT)** (Shishehchi and Banhashem, 2019): UAT is a questionnaire to assess the quality of job recommenders for people with disability, used by the authors in a user study. The questionnaire investigates four factors, i.e., usefulness, ease of use, ease of learning, and satisfaction with the RRS. The responses are registered in terms of values from 1 (disagree) to 5 (agree) for questions related to each factor.
- **Dataset Bias Metrics** (Delecraz et al., 2022): Statistical parity (SP) and Disparate impact (DI) metrics are used for measuring the bias in datasets with respect to binary attribute $A \in \{a_1, a_2\}$. Here, $\rho(u, x) = 1$ denotes that candidate u is qualified for a job post x randomly sampled from the set of job posts:

$$SP(u) = P(\rho(u, x) = 1 | A = a_1) - P(\rho(u, x) = 1 | A = a_2)$$

$$DI(u) = \frac{P(\rho(u, x) = 1 | A = a_1)}{P(\rho(u, x) = 1 | A = a_2)}$$

2.5. Investigating fairness of recommendations

A subset of reviewed articles analyzes fairness in RRSs. Tang et al. (2017) examines 17 million LinkedIn job listings spanning over 10 years and conducts a user study to analyze perceived stereotypes in these listings. The authors use the recruitment assistance services company Textio⁹ and Unitive (no longer operational) to get a list of gendered words and then, based on the weighted frequency of those words in a job listing, measure their "maleness" and "femaleness." They find that job listings perceived as overall male and the usage of gendered words, in general, have decreased over the years. These results could suggest that our society is moving toward more gender-appropriate language. They also conducted a user study where two of the questions are "While reading the job description, to what extent did you feel that the advertisement would attract more male or more female applicants?" and "If you were fully qualified to apply for a job like this, how likely is it that you would apply for this particular position?" They compare user responses with the "genderedness" of job listings measured earlier using the Mann-Whitney U (MWU) test and found that there is a low correlation between the gendered wording of job listings and perceived gender bias (attractiveness to female applicants). Instead, the perceived bias depends on preconceived notions like technology jobs are male jobs or lower wage jobs are female jobs. Additionally, the willingness to apply had a low

9 <https://textio.com/> (accessed May 2023).

correlation with perceived gender bias or the gendered wording of job listing.

Similarly, [Chen et al. \(2018\)](#) investigates gender bias on Indeed, Monster, and CareerBuilder resume searches. They use a regression model to measure IF and DP using MWU. The IF for the candidates occurring at ranks 30 – 50 in the recommendation list shows that men occupy higher ranks compared to women, which can seem counter-intuitive. For DP, this advantage is significant for multiple job titles (e.g., Truck Driver and Software Engineer).

The work of [Delecraz et al. \(2022\)](#) analyzes bias of different attributes like age, gender, geography, and education in their private dataset with disparate impact (DI) and statistical parity (SP). They found that their dataset is fair along gender and age attributes but is unfair according to nationality. They use true positive rate parity (TPRP) to analyze EO over their private candidate recommender. They found that education, birthplace, and residence permit were impactful for the candidate selection of individuals, while age and gender were not. This is partially explained by the fact that education, birthplace, and residence permit, in many cases, can be seen as requirements rather than biases.

[Elbassuoni et al. \(2019, 2020\)](#) propose new heuristic- and decision-tree-based approaches, respectively, to find a partition of candidates based on their attributes for which unfairness in terms of DP is maximized. After partitioning, the authors use EMD to measure the unfairness in the ranked list. Then, following a similar methodological approach for measuring bias, [Amer-Yahia et al. \(2020\)](#) investigate the online recruitment platform TaskRabbit and the job recommender platform Google Job Search. The uniqueness of [Elbassuoni et al. \(2019, 2020\)](#) and [Amer-Yahia et al. \(2020\)](#) is that the authors consider partitioning based on combinations of attributes rather than single attributes.

[Wilson et al. \(2021\)](#) conduct a fairness audit of the Pymetrics candidate screening system.¹⁰ Through this audit, the authors try to verify Pymetrics' claim to abide the 4/5th rule from the US Union Guidelines on Employee Selection Procedures ([Cascio and Aguinis, 2001](#)). According to the 4/5th rule, if the selection rate of a group is less than 4/5th of the highest group selection rate, then that group is adversely impacted. This rule closely aligns with the DP definition of fairness.

[Zhang \(2021\)](#) provides a gender fairness audit of four Chinese job boards. On these job boards, there are some job posts with explicit mention of preferred gender. The audit using list/set difference rate (LDR/SDR) showed the existence of gender bias in terms of quality of recommendation, and also differences in the wording of job ads recommended to males versus those recommended to female job seekers.

[Markert et al. \(2022\)](#) is the only article to pursue individual fairness (IF) and adapt the classification IF certification (IFC) process for ranking. The IF definition for candidate recommenders in Section 2.3 requires similar candidates to have similar ranks. A regression model is trained to predict a candidate's rank in the recommended list given by the candidate recommender. The authors formulate a mixed integer linear programming problem

using the regression model and the similarity constraint to get upper and lower bounds for the output of the regression model, i.e., the candidate's rank.

2.6. Pre-, in-, and post-processing approaches for fairness

The approaches to achieve fairness in RRSs can be categorized into pre-, in-, and post-processing techniques.

Pre-processing approaches are applied to the training data of the RS. The following approaches are used in the literature on RRSs:

- *Balancing the dataset*: Balancing the training data with respect to the sensitive attribute. For instance, [Arafan et al. \(2022\)](#) create gender-balanced synthetic data using CTGAN (conditional tabular generative adversarial network) for training ([Xu et al., 2019](#)), resulting in a significant decrease in NDKL.
- *Replacing the pronouns*: A simple approach for gender bias mitigation is to replace gendered pronouns with gender-neutral pronouns, as performed for instance in [Rus et al. \(2022\)](#) and [Jourdan et al. \(2023\)](#). This approach shows no change in terms of TPRP compared to not using pronoun substitution for [Rus et al. \(2022\)](#), while [Jourdan et al. \(2023\)](#) show a slight improvement in TPRP. The difference in results could be due to the different language models used.
- *Constrained resume sourcing*: [Syed and Shivendu \(2022\)](#) find conditions regarding the number of relevant candidates in each subgroup at the data sourcing of resumes for training to reduce the TRPR and theoretically achieve EO and DP.
- *Special group RSs*: The ontology-based/KB job recommenders use the sensitive information (e.g., disability, age, location, language) as input to create dedicated RSs for special groups (e.g., migrants or disabled people) ([Shishehchi and Banihashem, 2019](#); [Ntioudis et al., 2022](#)). [Shishehchi and Banihashem \(2019\)](#) show strong acceptance of their RS by the special group (disabled people) using UAT, while [Ntioudis et al. \(2022\)](#) do not evaluate their system for the special group (migrant people).

In-processing approaches to mitigate bias change the RS itself to make the recommendations less biased. The in-processing approach is the least used method in the RRS literature. Also, in-processing approaches are only used with CBF.

- *Adversarial debiasing*: [Rus et al. \(2022\)](#) fine-tune a large language model to learn job-candidate similarity with the additional objective of removing gender information from the embedding of job posts, using an adversarial network that tries to predict the gender of the candidate. The adversarial debiasing shows significant improvement in TPRP compared to replacing pronouns.
- *Regularization-based debiasing*: Methods such as [Jourdan et al. \(2023\)](#) use a regularization term in the loss. network. In this case, the regularization term is the Sinkhorn Divergence ([Chizat et al., 2020](#)) over the distribution of sensitive attributes in the predictions. Similar to adversarial debiasing, this

¹⁰ <https://www.pymetrics.ai/> (accessed May 2023).

approach shows significant improvement in TPRP compared to replacing pronouns.

Post-processing approaches are applied to the ranking received from the RRS to re-rank the items (jobs or candidates in our case). In research on fairness of RRSs, post-processing approaches are the most common, then pre-processing approaches, and in-processing are least explored in the literature (see Table 2), which is strikingly different from the fairness research in RSs overall (Deldjoo et al., 2023), where this order is in-, post-, and pre-processing, respectively. This comparison should be seen with caution though, as the number of papers surveyed is significantly less here compared to Deldjoo et al. (2023). The following are the post-processing approaches used in the RRSs we identified:

- *Introducing proportional fairness constraints*: Bei et al. (2020) try to achieve PF by removing the lowest ranking candidate inside the recommended list for which the attributes' upper bound condition (i.e., β) of PF gets violated, and adding the highest ranking candidate outside the recommended list for which the attributes' lower bound condition (i.e., α) of PF gets violated. This approach is able to approach PF with very few constraints violated.
- *Deterministic constrained sorting*: Geyik et al. (2019) introduce a deterministic sorting algorithm with constraints similar to the PF constraints. The difference with the work of Bei et al. (2020) is that here each candidate has only one attribute rather than a set of attributes and the size of the recommended list is not fixed here. Arafan et al. (2022) re-rank the recommended list to achieve the same number of candidates for each attribute using the re-ranking algorithm of Geyik et al. (2019). Both papers show improvement in NDKL. Arafan et al. (2022) additionally show that the NDKL scores of the deterministic constrained sorting approach can be further improved by using an artificially balanced dataset.
- *Spatial partitioning*: Burke et al. (2021) introduce a re-ranking algorithm based on spatial partitioning of 3-dimensional space created by three attributes of the candidate, i.e., performance, retention after hiring, and whether the candidate belongs to a protected group or not. Improvements in terms of AIR score are mentioned by Burke et al. (2021) though scores are not reported.
- *Targeting user group fairness*: Li et al. (2023) re-rank the recommended list of jobs for all candidates to maximize the sum of each user's personalized utility score (Zhang et al., 2016) over all candidate-job pairs while minimizing the UGF value.
- *Intervention-based skill improvement*: The method proposed by Scher et al. (2023) selects candidates to upgrade their skills so that it improves their probability of selection by the candidate recommenders. The authors divide the candidates into high-prospect group and low-prospect group using their skills and sensitive information as decision criteria. Then, the high-prospect groups' skill is upgraded and the skill upgrade of the low-prospect group is delayed for some time. The method helps the high-prospect group and punishes the low-prospect group, resulting in less reduction of inequality (i.e.,

less improvement in EO and DP) in the long run compared to random selection of candidates. And the selection of only low-prospect group has no impact on EO in the long run.

- *Attribute intervention*: Tran et al. (2021) and Tran et al. (2022) identify the skill set of the candidate that should be upgraded and the level to which it should be upgraded based on candidate attributes (including protected attributes) to achieve improvement in selection probability by a causal tree-based candidate recommender, more precisely, by a maximal causal tree (Tran et al., 2021) or personalized causal tree (Tran et al., 2022).

The reviewed debiasing approaches work toward fairer RRSs from a technical perspective, but the recruitment domain requires to take a legal perspective too since obligations and requirements from employment law apply. In the next section, we will therefore give an overview of the most critical legal pitfalls for RRSs. With this, we aim at giving the researchers and developers of RRSs guidelines on possible legal issues of their systems.

3. Legal validity

When RSs are used in the hiring process to (help) make decisions, the legal requirements from employment law are applicable to them in a similar manner as to human decision-makers (Barocas et al., 2019). In the case of hiring decisions, with or without algorithmic support, *non-discrimination law* (European Council, Directive 2000/43/EC, 2000; European Council, Directive 2000/78/EC, 2000; European Parliament and the Council, Directive 2006/54/EC, 2006) and *data protection law* [European Parliament and the Council, Regulation (EU) 2016/679, 2016] in particular must be observed (Hacker, 2018). RS users, in addition, should also take note of newly adopted laws regulating the online realm and AI technology in the EU.¹¹ Most notably in this context are the Digital Markets Act [European Parliament and the Council, Regulation (EU) 2022/1925, 2022] and the Digital Services Act [European Parliament and the Council, Regulation (EU) 2022/2065, 2022] which are already in force, as well as the Proposal for an Artificial Intelligence Act (European Commission, Proposal Artificial Intelligence Act, 2021) which could become law within 2024. While ordinarily the term "fairness" is used to describe fair and equal divisions of resources, in law the concept dealing with this in the world of employment is "non-discrimination." The fundamental right of non-discrimination is highlighted in all of the recent EU regulations. In the specific case of RRSs, by virtue of being applied in recruitment, the algorithmic decision has to comply with the directives of non-discrimination law.

Compared to these legal sources and terminologies, researchers and practitioners in computer science (CS) and artificial intelligence (AI) commonly use the term "fairness" and quantify it according to some computational metric, as we have seen in Sections 2.3 and 2.4. Therefore, their assumption is that a system can be more or less fair. In stark contrast, the legality of a

¹¹ <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> (accessed May 2023).

decision cannot be measured, it is either a case of discrimination or it is not. The benchmark is non-discrimination law. A system cannot in principle base its hiring decision on the sex, race or ethnicity, religion or belief, disability, age, or sexual orientation of the candidates, as EU law prohibits unjustified differential treatment on the basis of these protected characteristics (direct discrimination). A system can also not use an apparently neutral criterion which will have the effect of disadvantaging a considerably higher percentage of persons sharing the protected characteristic (indirect discrimination). This would happen if a system is applied in the same way to everybody, but disadvantages a group of people who share a protected characteristic.

The legal literature, despite the terminological incongruence, has picked up the most used fairness definitions in CS and AI research and often categorized algorithms that adopt them as either blind (or unaware) algorithms or protected-characteristics-aware algorithms (Žliobaitė and Custers, 2016; Bent, 2019; Wachter et al., 2020; Kim, 2022). The former describes algorithmic designs that achieve their results without using any of the protected characteristics as grounds for decision. The algorithm is, therefore, not given any information about any of the protected attributes. The latter describes algorithms which use some or all protected characteristics in their data to make decisions. Blind algorithms not only tend to underperform (Žliobaitė and Custers, 2016; Bent, 2019; Xiang and Raji, 2019; Wachter et al., 2020; Kim, 2022), but also often learn spurious correlations in the data that can serve as proxies for the protected characteristics in their datasets, which can equally lead to discrimination cases (Žliobaitė and Custers, 2016; Chander, 2017; Kim, 2017; Bent, 2019; Wachter et al., 2020; Adams-Prassl, 2022; Hildebrandt, 2022).¹² This can occur, for instance, when a correlation between ethnicity and postal code is drawn or when the recommendation algorithm unintentionally incorporates implicit gender information from interaction data because of different preferences of male and female users (Ganhör et al., 2022). The fairness definitions presented above all use protected characteristics to achieve fairness (Barocas et al., 2019; Bent, 2019).

3.1. Are debiased candidate recommenders affirmative action measures?

Debiased candidate recommenders essentially re-rank candidates according to the fairness definition used in order to achieve a parity for groups with and without certain protected characteristics. Job recommenders inherently also perform a re-ranking, however of job advertisement and employment opportunities. The discrimination risks thereby are similar, albeit center more on the delivery system (e.g., targeted ads). A closer investigation though transcends the scope of this survey (Greif and Grosz, 2023). Affirmative action schemes, known as “positive action” in EU law, are measures by which specific advantages are given to the underrepresented groups in order to compensate for existing disadvantages in working life to ensure full equality. Quotas are the most used example and directly relate to what RRSs

are doing by re-ranking candidates. “Fair” algorithmic affirmative action, as strived for by the fairness definitions, does however not translate to a lawful understanding of the concept (Xiang and Raji, 2019). The Court of Justice of the EU (CJEU) has previously stated that schemes which give “absolute and unconditional priority” exceed the limitations of the positive action exception (CJEU C-450/93, 1995). In subsequent case-law, the CJEU however narrowed the scope explaining that flexible quotas allowing for individual consideration would be in line with the positive action exception (CJEU C-409/95, 1997). As long as a “saving clause” is provided for, allowing for an objective assessment of all criteria, which can “override the priority accorded [...] where one or more of those criteria tilts the balance in favor of [another] candidate,” a quota scheme would be in line with EU law (CJEU C-409/95, 1997). Thus, every case involving the hiring according to a quota scheme must be open for individual consideration. Whether this could be computable is doubtful (Hacker, 2018; Adams-Prassl, 2022). Any candidate recommender re-ranking candidates to achieve a fixed fairness definition without taking each candidates’ individual circumstances into account will most likely run afoul of the legal requirements set out. For instance, imagine an employer deciding between two candidates, A and B, to fill one position. A and B are materially equally qualified, meaning that their resumes, though not necessarily the same, are of equivalent value for this position. Candidate A should be hired according to the affirmative action scheme in place. The employer however should (according to the “savings clause”) still hire candidate B, if there are individual circumstances, reasons specific to that candidate (e.g., sole provider or long term unemployed), which tilts the balance in candidate B’s favor. This needs to be decided on an individual case level and is therefore hard to automatize in a RRSs. This, however, must not mean the end of all algorithmically assisted hiring.

3.2. Pre-, in-, or post-processing: walking the legal line

As Hacker (2018) notes, the case-law dealing essentially with “corrective powers” at or after the selection process faces greater scrutiny from the CJEU than measures applied before. On the other hand, the CJEU seems to be more lenient when positive action measures are applied *before* first selection (CJEU C-158/97, 2000). Indeed, “positive measures” (even strict quotas) before the actual selection stage of the decision procedure are more likely to be accepted by the CJEU (Hacker, 2018). This could make implementing algorithmic fairness during the *training* stage of a model and before actually ranking candidates wrapped as a quota scheme widely applicable (Hacker, 2018; Adams-Prassl, 2022). Approaches including balancing the training dataset or re-ranking of (fictitious) candidates in the training phase of a candidate recommender should in principle be valid options for routing out biases before the model is put on the market. A glance over to the US shows a similar approach: US case-law suggests that rearrangement in terms of affirmative action applied *after* the selection results have been allocated to the respective

¹² Proxy discrimination most often comes in the form of indirect discrimination.

candidates¹³ could lead to discrimination of the now down-ranked selected candidates (US Supreme Court, *Ricci v DeStefano*, 2009).

For both jurisdictions (EU and US), the crux is timing: taking into account biases (and potentially discrimination) that is found in society and in datasets is possible (also via algorithmic help), as long as it is done *before* real-world application. Otherwise, a candidate recommender risks producing further discrimination by ranking candidates first and foremost according to a protected characteristic (e.g., sex) and not according to actual suitability for the job in question. Kim (2022) notes additionally that the US Supreme Court acknowledges that an employer may need to take protected characteristics into account to create fairer hiring processes. Along with Bent (2019), Kim (2017) argues that this should leave room for algorithmic affirmative action before the selection process. Therefore, as it stands, in terms of (both EU and US) non-discrimination law, pre-processing approaches should be favored, whereas post- and probably in-processing approaches would most likely run afoul of current requirements.

4. Conclusions and open challenges

In this survey, we analyzed the current fairness research in RRSs from multiple perspectives. First, the algorithms used in RRSs were classified into four categories. Subsequently, we consolidated different fairness definitions in the existing literature to understand the objectives of fairness research in RRSs. We provided the fairness definitions used in classification and connected them with their adapted forms for RRSs. Further, we detailed some of the fairness metrics found in the surveyed literature. We also discussed the work done to analyze the presence or absence of fairness in RRSs. Subsequently, the most common pre-, in-, and post-processing approaches to gain fairness in RRSs were described. Finally, we bridge the gap to legal scholarship by discussing fairness definitions and their relation to legal requirements in order to provide an overview of some of the possible legal issues resulting from unfairness in RRSs. We thereby identified the lack of interdisciplinary vocabulary and understanding as a substantial challenge.

RRSs is a quickly evolving field but is also facing several challenges and open questions which are waiting to be solved. First, the *lack of public datasets for fairness research* in the recruitment domain is evident, as highlighted in Table 2. The usage of private data and proprietary RRSs limits the understanding and reproducibility of fairness research.

Experiments are most often *limited to gender* as sensitive attribute. Other attributes such as ethnicity, age, or disability are commonly only targeted when studied in conjunction with gender and are often modeled artificially. This limits our understanding of the role of non-gender attributes in the recruitment domain. At the same time, it is exciting to see the recent works considering discrimination against groups

defined by not only a single attribute but a combination of attributes.

Fairness definitions are also an increasing challenge for the community. They are most commonly *adapted from classification fairness definitions* for RRSs. The adapted definitions for job recommenders does not consider the ranks of recommended items. This, however, is important for both candidate and job recommenders. Individual fairness (Markert et al., 2022) is an exciting new direction away from the standard of group-associated fairness definitions such as demographic parity and equal opportunity.

The fairness metrics are highly diverse across the literature surveyed. We would like to highlight here that even when the fairness definitions targeted are the same, the metrics used for fairness measurement are rarely identical. This points to the problem of a *lack of standardized fairness evaluation metrics* in the recruitment domain.

The recruitment domain is a content-rich domain (i.e., resumes and job posts both convey lots of descriptive textual semantics), which explains the prevalence of CBF and KB recommendation algorithms. The *scarcity of hybrid and CF algorithms* in RRSs research on fairness shows a disconnect with common RRSs research (de Ruijt and Bhulai, 2021). In addition, while currently *post-processing approaches are most often adopted* for debiasing, as discussed in Sections 2.6, 3.2, they are also the most concerning ones from a legal perspective. We, therefore, suggest to devote more research to pre-processing and in-processing strategies. In addition, we strongly advocate for more interdisciplinary research, involving experts from both RRSs and legal scholarship, to formulate strategies and constraints for legally suitable in- and post-processing approaches and to drive RS research in the recruitment domain.

To wrap up, the surveyed research works are (1) diverse in terms of job/candidate recommenders and the adopted algorithms, (2) explore new fairness definitions such as IF, and (3) experiment with various metrics that are attempts to better represent the same fairness concept. The current trajectory of fairness research in RRSs is highly promising, but several avenues for further improvements through the valuable input from diverse research communities is required.

Author contributions

Conceptualization and funding acquisition: MS, EG, and NR. Literature search and writing—original draft: DK and TG. Writing—review and editing: MS, NG, and EG. All authors approved the submitted version.

Funding

This research was funded by the Austrian Science Fund (FWF): DFH-23 and P33526 and by the State of Upper Austria and the

¹³ In the case of RRSs, this means that candidates have received their score from the system, based on which the ranked list is created.

Federal Ministry of Education, Science, and Research, through grants LIT-2020-9-SEE-113 and LIT-2021-YOU-215.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abel, F., Benczúr, A., Kohlsdorf, D., Larson, M., and Pálovics, R. (2016). "Recsys challenge 2016: Job recommendations," in *Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16* (New York, NY, USA: Association for Computing Machinery), 425–426. doi: 10.1145/2959100.2959207
- Abel, F., Deldjoo, Y., Elahi, M., and Kohlsdorf, D. (2017). "Recsys challenge 2017: Offline and online evaluation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17* (New York, NY, USA: Association for Computing Machinery), 372–373. doi: 10.1145/3109859.3109954
- Adams-Prassl, J. (2022). Regulating algorithms at work: Lessons for a "European approach to artificial intelligence." *Eur. Labour Law J.* 13, 30–50. doi: 10.1177/20319525211062558
- Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proc. ACM Hum.-Comput. Interact.* 3, 1–30. doi: 10.1145/3359301
- Amer-Yahia, S., Elbassuoni, S., Ghizzawi, A., Borromeo, R. M., Hoareau, E., and Mulhem, P. (2020). "Fairness in online jobs: A case study on TaskRabbit and Google," in *International Conference on Extending Database Technologies (EDBT)* (Copenhagen, Denmark).
- Arafan, A. M., Graus, D., Santos, F. P., and Beauxis-Aussalet, E. (2022). "End-to-end bias mitigation in candidate recommender systems with fairness gates," in *2nd Workshop on Recommender Systems for Human Resources, RecSys-in-HR 2022*. CEUR-WS.
- Barocas, S., Hardt, M., and Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA, MIT press.
- Becker, B., and Kohavi, R. (1996). Adult. *UCI Machine Learning Repository*. Available online at: <https://archive-beta.ics.uci.edu/dataset/2/adult> (accessed September 27, 2023).
- Bei, X., Liu, S., Poon, C. K., and Wang, H. (2020). "Candidate selections with proportional fairness constraints," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 150–158.
- Bent, J. R. (2019). Is algorithmic affirmative action legal. *Geo. LJ* 108, 803.
- Bian, S., Chen, X., Zhao, W. X., Zhou, K., Hou, Y., Song, Y., et al. (2020). "Learning to match jobs with resumes from sparse interaction data using multi-view co-teaching network," in *Proceedings of the 29th ACM International Conference on Information Knowledge Management, CIKM '20* (New York, NY, USA: Association for Computing Machinery), 65–74. doi: 10.1145/3340531.3411929
- Burke, I., Burke, R., and Kuljanin, G. (2021). "Fair candidate ranking with spatial partitioning: Lessons from the SIOP ML competition," in *Proceedings of the First Workshop on Recommender Systems for Human Resources (RecSys in HR 2021) co-located with the 15th ACM Conference on Recommender Systems (RecSys 2021)* (CEUR-WS).
- Cascio, W. F., and Aguinis, H. (2001). The federal uniform guidelines on employee selection procedures (1978), an update on selected issues. *Rev. Public Person. Adm.* 21, 200–218. doi: 10.1177/0734371X0102100303
- Chander, A. (2017). The racist algorithm? *Michigan Law Rev.* 115, 1023. doi: 10.36644/mlr.115.6.racist
- Chen, L., Ma, R., Hannák, A., and Wilson, C. (2018). "Investigating the impact of gender on rank in resume search engines," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY, USA: Association for Computing Machinery), 1–14. doi: 10.1145/3173574.3174225
- Chizat, L., Roussillon, P., Léger, F., Vialard, F.-X., and Peyré, G. (2020). "Faster Wasserstein distance estimation with the Sinkhorn divergence," in *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20* (Red Hook, NY, USA: Curran Associates Inc.).
- CJEU C-158/97 (2000). Georg Badeck and Others v Landesanwalt beim Staatsgerichtshof des Landes Hessen.
- CJEU C-409/95 (1997). Hellmut Marschall v Land Nordrhein-Westfalen.
- CJEU C-450/93 (1995). Eckhard Kalanke v Freie Hansestadt Bremen.
- Corder, G. W., and Foreman, D. I. (2014). *Nonparametric Statistics: A Step-by-Step Approach*. New York, NY: John Wiley Sons.
- Dastin, J. (2018). "Amazon scraps secret AI recruiting tool that showed bias against women," in *Ethics of Data and Analytics* (Auerbach Publications), 296–299. doi: 10.1201/9781003278290-44
- de Ruijt, C., and Bhulai, S. (2021). Job recommender systems: A review. *arXiv preprint arXiv:2111.13576*.
- De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., et al. (2019). "Bias in bios: A case study of semantic representation bias in a high-stakes setting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19* (New York, NY, USA: Association for Computing Machinery), 120–128. doi: 10.1145/3287560.3287572
- Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., and Zanzonelli, D. (2023). Fairness in recommender systems: Research landscape and future directions. *User Model User-Adap. Inter.* 30, 1–50. doi: 10.1007/s11257-023-09364-z
- Delecraz, S., Eltarr, L., Becuwe, M., Bouxin, H., Boutin, N., and Oullier, O. (2022). "Making recruitment more inclusive: Unfairness monitoring with a job matching machine-learning algorithm," in *Proceedings of the 2nd International Workshop on Equitable Data and Technology, FairWare '22* (New York, NY, USA: Association for Computing Machinery), 34–41. doi: 10.1145/3524491.3527309
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Minneapolis, MA: Association for Computational Linguistics), 4171–4186.
- Dhameliya, J., and Desai, N. (2019). "Job recommender systems: A survey," in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)* 1–5. doi: 10.1109/i-PACT44901.2019.8960231
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness through awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS '12* (New York, NY, USA: Association for Computing Machinery), 214–226. doi: 10.1145/2090236.2090255
- Elbassuoni, S., Amer-Yahia, S., and Ghizzawi, A. (2020). Fairness of scoring in online job marketplaces. *ACM/IMS Trans. Data Sci.* 1, 1–30. doi: 10.1145/3402883
- Elbassuoni, S., Amer-Yahia, S., Ghizzawi, A., and Atie, C. E. (2019). "Exploring fairness of ranking in online job marketplaces," in *22nd International Conference on Extending Database Technology (EDBT)* (Lisbone, Portugal).
- European Commission, Proposal Artificial Intelligence Act (2021). Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts: COM 2021 206 final.
- European Council, Directive 2000/43/EC (2000). Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin: Race Directive.
- European Council, Directive 2000/78/EC (2000). Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation: Framework Directive.
- European Parliament and the Council, Directive 2006/54/EC (2006). Directive 2006/54/EC on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast): Gender Recast Directive.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- European Parliament and the Council, Regulation (EU) 2016/679 (2016). The Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive (EC) 95/46 (General Data Protection Regulation): GDPR.
- European Parliament and the Council, Regulation (EU) 2022/1925 (2022). Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act): DMA.
- European Parliament and the Council, Regulation (EU) 2022/2065 (2022). Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive (EC) 2000/31 (Digital Services Act): DSA.
- Fawaz, N. (2019). Recommending dream jobs in a biased real world. *arXiv preprint arXiv:1905.06134*.
- Freire, M. N., and de Castro, L. N. (2021). E-recruitment recommender systems: A systematic review. *Knowl. Inf. Syst.* 63, 1–20. doi: 10.1007/s10115-020-01522-8
- Ganhör, C., Penz, D., Rekasabs, N., Lesota, O., and Schedl, M. (2022). “Unlearning protected user attributes in recommendations with adversarial training,” in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22* (New York, NY, USA: Association for Computing Machinery), 2142–2147.
- Geyik, S. C., Ambler, S., and Kenthapadi, K. (2019). “Fairness-aware ranking in search recommendation systems with application to LinkedIn talent search,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, KDD '19* (New York, NY, USA: Association for Computing Machinery), 2221–2231. doi: 10.1145/3292500.3330691
- Greif, E., and Grosz, T. (2023). To see, or Not to see: Online job advertisement and EU non-discrimination law. *Eur. Labour Law J.* 14, 20319525231172089. doi: 10.1177/20319525231172089
- Gugnani, A., and Misra, H. (2020). “Implicit skills extraction using document embedding and its use in job recommendation,” in *Proceedings of the AAAI Conference on Artificial Intelligence* 13286–13293. doi: 10.1609/aaai.v34i08.7038
- Gutiérrez, F., Charleer, S., De Croon, R., Htun, N. N., Goetschalckx, G., and Verbert, K. (2019). “Explaining and exploring job recommendations: A user-driven approach for interacting with knowledge-based job recommender systems,” in *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY, USA: Association for Computing Machinery), 60–68. doi: 10.1145/3298689.3347001
- Hacker, P. (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common. Market Law Rev.* 55, 1143–1186. doi: 10.54648/COLA2018095
- Hardt, M., Price, E., Price, E., and Srebro, N. (2016). “Equality of opportunity in supervised learning,” in *Advances in Neural Information Processing Systems*, eds. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Red Hook, NY: Curran Associates, Inc.).
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). “Neural collaborative filtering,” in *Proceedings of the 26th International Conference on World Wide Web, WWW '17* (Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee), 173–182. doi: 10.1145/3038912.3052569
- Hildebrandt, M. (2022). The Issue of Proxies and Choice Architectures. Why EU law matters for recommender systems. *Front. Artif. Intell.* 73, 789076. doi: 10.3389/frai.2022.789076
- Jourdan, F., Kaninku, T. T., Asher, N., Loubes, J.-M., and Risser, L. (2023). How optimal transport can tackle gender biases in multi-class neural network classifiers for job recommendations. *Algorithms* 16, 174. doi: 10.3390/a16030174
- Kim, P. T. (2017). Data-driven discrimination at work. *Wm. Mary L. Rev.* 58, 857.
- Kim, P. T. (2022). Race-aware algorithms: Fairness, nondiscrimination and affirmative action. *Cal. L. Rev.* 110, 1539. doi: 10.15779/Z387P8TF1W
- Koenig, N., and Thompson, I. (2021). “The 2020–2021 SIOP machine learning competition,” in *36th annual Society for Industrial and Organizational Psychology Conference* (New Orleans, LA: SIOP).
- Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer* 42, 30–37. doi: 10.1109/MC.2009.263
- Lacic, E., Reiter-Haas, M., Kowald, D., Reddy Daredy, M., Cho, J., and Lex, E. (2020). Using autoencoders for session-based job recommendations. *User Model. User-Adapt. Inter.* 30, 617–658. doi: 10.1007/s11257-020-09269-1
- Li, Y., Yamashita, M., Chen, H., Lee, D., and Zhang, Y. (2023). “Fairness in job recommendation under quantity constraints,” in *AAAI-23 Workshop on AI for Web Advertising*.
- Liu, Q., Zeng, Y., Mokhosi, R., and Zhang, H. (2018). “STAMP: Short-term attention/memory priority model for session-based recommendation,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, KDD '18* (New York, NY, USA: Association for Computing Machinery), 1831–1839. doi: 10.1145/3219819.3219950
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *CoRR, abs/1907.11692*.
- Luo, Y., Zhang, H., Wen, Y., and Zhang, X. (2019). “ResumeGAN: An optimized deep representation learning framework for talent-job fit via adversarial learning,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19* (New York, NY, USA: Association for Computing Machinery), 1101–1110. doi: 10.1145/3357384.3357899
- Markert, K., Ahouzi, A., and Debus, P. (2022). Fairness in regression-analyzing a job candidates ranking system. *INFORMATIK* 2022, 1275–1285. doi: 10.18420/inf2022_109
- Ntioudis, D., Masa, P., Karakostas, A., Meditskos, G., Vrochidis, S., and Kompatsiaris, I. (2022). Ontology-based personalized job recommendation framework for migrants and refugees. *Big Data Cogn. Comput.* 6, 120. doi: 10.3390/bdcc6040120
- Pele, O., and Werman, M. (2009). “Fast and robust earth mover’s distances,” in *2009 IEEE 12th International Conference on Computer Vision* 460–467. doi: 10.1109/ICCV.2009.5459199
- Rafter, R., Bradley, K., and Smyth, B. (2000). “Personalised retrieval for online recruitment services,” in *The BCS/IRSG 22nd Annual Colloquium on Information Retrieval (IRSG 2000)* (Cambridge, UK).
- Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). “Mitigating bias in algorithmic hiring: Evaluating claims and practices,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20* (New York, NY, USA: Association for Computing Machinery), 469–481. doi: 10.1145/3351095.3372828
- Ruoss, A., Balunović, M., Fischer, M., and Vechev, M. (2020). “Learning certified individually fair representations,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20* (Red Hook, NY, USA: Curran Associates Inc.).
- Rus, C., Luppens, J., Oosterhuis, H., and Schoenmacker, G. H. (2022). “Closing the gender wage gap: Adversarial fairness in job recommendation,” in *2nd Workshop on Recommender Systems for Human Resources, RecSys-in-HR 2022. CEUR-WS*.
- Salakhutdinov, R., and Mnih, A. (2007). “Probabilistic matrix factorization,” in *Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07* (Red Hook, NY, USA: Curran Associates Inc.), 1257–1264.
- Scher, S., Kopeinik, S., Trügler, A., and Kowald, D. (2023). Modelling the long-term fairness dynamics of data-driven targeted help on job seekers. *Scient. Rep.* 13, 1727. doi: 10.1038/s41598-023-28874-9
- Shishchichi, S., and Banihashem, S. Y. (2019). JRDP: A job recommender system based on ontology for disabled people. *Int. J. Technol. Hum. Interact.* 15, 85–99. doi: 10.4018/IJTHI.2019010106
- Smith, S. M., Edwards, R., and Duong, H. C. (2021). Unemployment rises in 2020, as the country battles the COVID-19 pandemic. *Monthly Lab. Rev.* 144, 1. doi: 10.21916/mlr.2021.12
- Syed, R. A., and Shivendu, S. (2022). *Sourcing bias on job matching platforms: Regulation and optimal platform strategy*. Available online at: <https://ssrn.com/abstract=4188162> (accessed June 15, 2022). doi: 10.2139/ssrn.4188162
- Tang, S., Zhang, X., Cryan, J., Metzger, M. J., Zheng, H., and Zhao, B. Y. (2017). Gender bias in the job market: A longitudinal analysis. *Proc. ACM Hum.-Comput. Interact.* 1, 1–19. doi: 10.1145/3134734
- Thali, R., Mayekar, S., More, S., Barhate, S., and Selvan, S. (2023). “Survey on job recommendation systems using machine learning,” in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* 453–457. doi: 10.1109/ICIDCA56705.2023.10100122
- Tran, H. X., Le, T. D., Li, J., Liu, L., Liu, J., Zhao, Y., et al. (2021). “Recommending the most effective intervention to improve employment for job seekers with disability,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining, KDD '21* (New York, NY, USA: Association for Computing Machinery), 3616–3626. doi: 10.1145/3447548.3467095
- Tran, H. X., Le, T. D., Li, J., Liu, L., Liu, J., Zhao, Y., et al. (2022). “Recommending personalized interventions to increase employability of disabled jobseekers,” in *Advances in Knowledge Discovery and Data Mining: 26th Pacific-Asia Conference, PAKDD 2022, Chengdu, China, May 16–19, 2022, Proceedings, Part III* (Berlin, Heidelberg: Springer-Verlag), 92–104. doi: 10.1007/978-3-031-05981-0_8
- US Supreme Court. (2009). *Ricci v DeStefano*. 557 U.S. 557 2009.
- Vega, J. (1990). “Semantic matching between job offers and job search requests,” in *COLING 1990 Volume 1: Papers presented to the 13th International Conference on Computational Linguistics*. doi: 10.3115/992507.992520
- Wachter, S., Mittelstadt, B., and Russell, C. (2020). Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. *West Virginia Law Rev.* 123, 735. doi: 10.2139/ssrn.3792772
- Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.* 41, 1–43. doi: 10.1145/3594633
- Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., et al. (2021). “Building and auditing fair algorithms: A case study in candidate screening,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21* (New York, NY, USA: Association for Computing Machinery), 666–677. doi: 10.1145/3442188.3445928
- Wilson, R., Beaven, R., May-Gillings, M., Hay, G., and Stevens, J. (2014). *Working Futures 2012–2022*. Evidence Report 83.
- Xiang, A., and Raji, I. D. (2019). On the legal compatibility of fairness definitions. *arXiv preprint arXiv:1912.00761*. doi: 10.48550/arXiv.1912.00761

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, eds. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Red Hook, NY, USA: Curran Associates, Inc.).

Zehlike, M., Yang, K., and Stoyanovich, J. (2022). Fairness in ranking, Part II: Learning-to-rank and recommender systems. *ACM Comput. Surv.* 55, 1–41. doi: 10.1145/3533380

Zhang, S. (2021). *Measuring algorithmic bias in job recommender systems: An audit study approach*. Technical report. doi: 10.1257/rct.6101-1.1

Zhang, Y., Zhao, Q., Zhang, Y., Friedman, D., Zhang, M., Liu, Y., et al. (2016). "Economic recommendation with surplus maximization," in *Proceedings of the 25th International Conference on World Wide Web, WWW '16* (Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee), 73–83. doi: 10.1145/2872427.2882973

Žliobaitė, I., and Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artif. Intell. Law* 24, 183–201. doi: 10.1007/s10506-016-9182-5