

Ethics Guidelines for Using AI-based Algorithms in Recruiting: Learnings from a Systematic Literature Review

Lennart Hofeditz
University of
Duisburg-Essen,
Germany
lennart.hofeditz@uni-due.de

Milad Mirbabaie
Paderborn
University,
Germany
milad.mirbabaie@uni-paderborn.de

Audrey Luther
University of
Duisburg-Essen,
Germany
audrey.luther@stud.uni-due.de

Riccarda Mauth
University of
Duisburg-Essen,
Germany
Riccarda.Mauth@stud.uni-due.de

Ina Rentemeister
University of
Duisburg-Essen,
Germany
ina.rentemeister@stud.uni-due.de

Abstract

To reduce the workload of employees working in Human Resource departments and to avoid bias in pre-selection of applicants, an increasing number of companies deploy Artificial Intelligence (AI)-based algorithms. Some examples such as Amazon’s discriminating recruiting algorithm showed that algorithms are not free of unethical decision making. Although there already exists a variety of ethics principles for AI-based systems, those are usually hardly being applicable to specific use cases such as using AI-based algorithms in recruiting processes. To address this issue and to provide guidance for researchers and practitioners, we conducted a systematic literature review (keyword and backwards search) on existing ethics guidelines and principles for AI and extracted aspects that seemed applicable to guide recruiting processes. Based on 28 relevant papers we derived actionable guidelines for using AI-based algorithms in recruiting processes. We categorized our guidelines into the aspects of fairness, avoidance of discrimination and avoidance of bias.

However, this is accompanied by many ethical concerns as we know from decades of research. Some employers have tended to discriminate against women and ethnic minorities, albeit unconsciously [7, 8, 9, 10]. Contrary to popular belief, discrimination occurs not only in traditional recruiting, but also in recruiting algorithms. Along with this there has been rapidly growing interest in the use of recruiting algorithms at least to address or even mitigate bias [6]. However, even if the developer of the algorithm does not intend to discriminate, the algorithm may contain a discriminating bias. One of the most prominent examples for a biased recruiting algorithm is the Amazon scandal of 2018. Build on historical job data the algorithm gave higher scores to white male applicants [13]. Even after the variable “gender” was removed, the algorithm learned to use other measures as a proxy for gender. This bias was introduced with the training data which reflected a decade-long human bias in the recruiting decisions of the company [14].

Current advances in research, development, and application of AI systems have yielded a far-reaching discourse on AI ethics. In consequence, several ethics guidelines have been released in recent years which comprise normative principles and recommendations. These guidelines were published from technological companies e.g., Microsoft, Google or IBM as well as from governmental institutions e.g., EU Ethics guidelines for trustworthy AI. However, a comprehensive systematic literature review that provide structure for these guidelines for research and practice in the context of recruiting is missing as existing systematic approaches in this field are limited to certain journals [15].

Furthermore, these guidelines run into the problem that they are either too generic to be applied in actual practice, too narrowly tailored to one area of use or concerned with the technical implementation or acceptance criteria of AI systems for recruiting [16].

1. Introduction

Artificial Intelligence (AI) in recruiting have evolved very quickly over the last decade [1, 2] Recruiting 3.0 refers to the hiring process enhanced by AI [1]. In 2019, a report estimated that about 99% of the fortune 500 companies use algorithms in their recruiting process [3]. The technical progress enables companies to make recruiting decisions supported by algorithms to save costs and achieve higher effectiveness [4, 5]. AI-based algorithms can support recruiting processed by analyzing applicants’ questionnaires, video interviews or CV data to perform a pre-selection for hiring candidates for a certain position [6].

In addition, an agenda for further research in this area would be highly valuable. Thus, we found that there is still no effective solution to the problem of unethical recruiting algorithms. It is inevitable to take a closer look at how ethical considerations and guidelines can influence the utilitarian towards a more moralistic perspective of the algorithm development.

Since negative aspects of recruiting algorithms, such as discrimination, can also be addressed with guidelines concerned with fairness, our research will analyze existing literature concerned with ethics guidelines for AI and will bridge the gap between those guidelines and AI recruiting algorithms. We analyzed existing guidelines on AI and ethics and addressed the following research question (RQ):

How can ethics principles for AI be applied to AI-based recruiting algorithms?

The aim of our research is to contribute to the development of ethics guidelines for recruiting algorithms by systematically reviewing existing literature on ethics and AI in recruiting as well as generally published ethics guidelines. Since Webster and Watson [17] and Liu, Gao and Mao [18] pointed out that emerging issues profit from literature reviews and recruiting algorithms are a phenomenon that is quickly evolving, there is the need for a systematic overview of the literature on ethics guidelines for AI in this context. We specified criteria that are essential in the development of ethics guidelines for recruiting algorithms and provided an overview of how ethic guidelines contribute to these criteria. Additionally, we propose specific guidelines for text analysis algorithms in the recruiting context.

2. Literature Background

2.1 AI and Recruiting

AI-based algorithms such as many supervised machine learning algorithms, are commonly used as standardization of established workplace decision [19]. Google, IBM, SAP and Microsoft facilitate Human Resources (HR) practices by offering algorithmic systems for recruiting and performance measures [20].

Algorithmic decision-making is used in HR and development in various large companies such as Intel, Vodafone or Ikea [21]. These algorithms provide the competitive advantage to review large numbers of applications automatically [1].

Algorithmic decision making has benefits like savings in time and cost, minimized risks, and increased productivity. Recently, certainty has been increasing in decision-making algorithms, making them even better suited for economic use [22, 24].

Besides these advantages, human biases like stereotypes and prejudices could be diminished by enhancing objectivity and fairness in recruiting decisions through decision-making algorithms [25]. In a demand for more objectivity, data-driven and evidence-based decision-making are used to combat unjust processes and bias in human decision [26]. However, multiple researchers point out social, legal, and ethical risks associated with algorithmic decision-making [27, 28].

2.2 Ethical Issues of AI in Recruiting

Research showed that processes using algorithmic decision making can result in negative outcomes for groups that have historically been disadvantaged [27]. Algorithms bear risks since they reflect the bias of the data scientists working on it and can magnify unnoticed patterns of discrimination [29]. The threat of discrimination, bias and unfairness is very likely to appear when working with and solely relying on algorithmic decision-making [30]. Trained on inaccurate, biased, or unrepresentative data, the outcomes produced by algorithms are discriminating and biased [30, 31].

The most frequent biases are gender and race bias which means having disadvantages because of gender or race [30, 31]. This can be assigned to discrimination, defined as the unequal treatment of different groups, not based on qualities but rather on gender, age or ethnicity [32]. Additionally, hidden and unintended biases are also dangerous [33]. When systems should be neutral and objective but are not, it is usually because the underlying algorithm lacks transparency, intelligibility and relies on training data which is biased in ways difficult to uncover [35]. Gender and race biases have been reported in decision-making algorithms in systems for many applications, including recruitment [34, 35, 31]. Often, this issue is only noticeable after the algorithm has already made the decision [36].

Numerous issues have been presented regarding decision-making algorithms in HR processes such as recruiting. HR outcomes are often very complex, making it hard to measure the dimensions of being a good employee, for example. From individual to team performance, widely-used biased measures like performance appraisal scores or the interdependency of a complex job to other jobs, there is uncertainty in the quality of performance evaluations [37, 38, 39].

The last problem raised by the authors is the one of implications of HR decisions, namely hiring and firing, since these can have “serious consequences for individuals and society with regard to ethics as well as to both procedural and distributive justice fairness”

[13, p.18]. Even though there are legal frameworks which hold the employer accountable in making fair decisions, state-of-the-art algorithms typically lack the aspect of explainability which helps understand the mechanisms and reasons for the decision of the algorithm [13].

These risks are especially apparent when a well-known company like Amazon stops using their discriminating recruiting algorithm after failing to implement changes making the decision-making algorithm suitable and non-discriminatory [14].

2.3 Ethics Guidelines for AI in Organizations

To counteract previously mentioned risks of AI, Asimov was the first author in 1940 to discuss guidelines for intelligent systems, the “Three Laws for Robotics”: namely that a robot may not injure a human being or allow a human being to come to harm, robots having to obey to human orders except in cases where it conflicts the first law and the last one being that a robot may protect his existence provided this does not conflict with the first two laws. While these laws were published in a fictional story, research has moved to real-world AI implementations and pointed out the need for ethical guidelines for data analytics, ML and AI [40].

In recent years, a whole body of guidelines has been developed by national and international organizations full of principles for companies and technology developers to adhere to [41]. Many more expert committees on AI ethics have appeared and drafted reports and guidelines on ethics in AI. While these are positive developments, the self-commitments by companies formulating their own ethical guidelines discourage the development of a binding legal framework [42].

Especially accountability, privacy, and fairness make-up around 80% of all guidelines, found by a review by Hagendorff [43], making those three aspects seem like the “[...] minimal requirements for building and using an “ethically sound” AI system” (p.103). There are many guidelines addressing these, among other, issues very generally or applied to a certain context. However, none of them address the issue of recruiting algorithms. Considering that this application is already implemented and has been for a few years, there is a need for comprehensive ethics guidelines in this area [44, 45]. This is where our research will offer new insights.

3. Method

We addressed our research question by conducting a systematic literature review (SLR)

according to Vom Brocke et al. [46] which further developed the classical approach of Webster and Watson [16]. We decided to conduct an SLR as an initial search of ethical guidelines for AI suggested that there already is a huge amount of literature regarding ethical guidelines for AI that could serve as a valuable basis for guidelines for recruiting algorithms. The systematic literature review “address[es] a specific research question that guides the data collection, extraction, and aggregation process” [46, p. 207], to ensure that only necessary papers regarding the research question are included in the analysis process. For The purpose of the review was a structured overall view of AI ethics guidelines for recruiting algorithms to synthesize existing research and provide avenues for future research according to Webster and Watson [17].

Before starting the search process, we defined a search scope to guarantee a structured and organized search [48]. The specification of the process according to Vom Brocke et al. [46] contained four steps. In the first step the process was determined. The following literature review was organized iteratively with an intertwined proceeding of reading, interpreting, and analyzing the identified literature [46]. Secondly, publications were specified as the source. In the third step, we applied inclusion and exclusion criteria according to Li, Gao and Mao [18]. As it was not purposeful to review the complete literature on ethics guidelines, we selected representative and seminal works [46] which have been published in peer-reviewed journals or conferences selected by our inclusion criteria. In the last step we defined our search technique, which was a keyword search combined with backward search according to Vom Brocke et al. [46]. The literature search was completed between January 22nd and February 5th, 2021. We scanned the following databases: IEEE, Scopus, AiSeL, ACM and Springer Link. In order to identify literature that could inform ethical guidelines for AI-based algorithms in recruiting, we used the following keywords as search strings: (Trust OR Distrust) AND (Recruitment AND (Artificial Intelligence)), ((Artificial Intelligence) AND (Ethic Guidelines)) AND Recruitment, (Artificial Intelligence) NEAR Ethics, (Artificial Intelligence) AND Discrimination OR Disadvantages.

During the search process it was recognized that the search terms were too narrow and focused on the term “Recruiting”. We therefore added six keywords in the further course that paraphrase the word “Recruiting”. These included: Trust AND Hiring AND (Artificial Intelligence), Distrust AND Hiring AND (Artificial Intelligence), (Artificial Intelligence) AND Ethic, Guidelines AND Ethics AND (Artificial

Intelligence), (Artificial Intelligence) AND Bias AND (Algorithmic Hiring).

We further examined title and abstract of each search result. We defined inclusion criteria and included articles where the content was not assigned to a specific discipline (such as healthcare), but discussed ethical guidelines, AI and recruiting in general. Based on the results of the keyword search, a backward search was completed. For this, we coded the identified papers through three independent coders. A total of 49 relevant articles resulted from this process.

In a second selection process the articles were screened in greater detail. Our aim was to identify several criteria that could be considered important regarding ethical guidelines that prevent unethical algorithms in recruiting. We excluded 21 paper due to four aspects: First, some of the results were too focused on the technical part of AI, resulting in that no relevant criteria could be identified. For this reason, we excluded two papers. Another exclusion criterion was that some did not really focus on the topic of recruiting which is why four papers were not further analyzed in the literature review. Third, in 13 articles no relevant criteria regarding the research question could be identified. After this filtering process, 28 relevant papers were included and further analyzed.

4. Results

4.1 Existing Ethical Principles for Using Algorithms in Recruiting

After reviewing the literature for the inclusion and exclusion criteria, we first inductively categorized the 28 articles into types of principles. We found three principles for AI-based recruiting algorithms that were mentioned most frequently: fairness, prevention of discrimination and avoidance of bias. We summarized our categorization in Table 1.

Table 1. Criteria for recruiting algorithms

Fairness
Hajian et al., 2016 Köchling & Wehner, 2020 Köchling et al. 2021 Ochmann & Laumer, 2019 Sánchez-Monedero et al., 2020 Xu et al., 2020 Zhou, 2019
Prevention of Discrimination

Fernández-Martínez & Fernández, 2020 Hajian et al., 2016 Köchling & Wehner, 2020 Kuleshov & Abramova., 2020 Raghavan et al., 2020
Avoidance of Bias
Xu et al., 2020 Köchling & Wehner, 2020 Ghallab, 2019 Kuleshov & Abramova, 2020

Köchling & Wehner [15]. stated, in HR recruitment and development there are two fairness types that need to be considered: Objective and subjective fairness perceptions of applicants and employees about the usage of algorithmic recruiting decisions. For example, during the recruitment process, perceptions of fairness can have consequences such as applicant's intent to seek employment or to provide referrals to other applicants and thus important impact for the decision to stay in the applicant pool or accept a job [15].

Ochmann & Laumer [47] assume that fairness is a construct that encompasses both the legal framework and the consideration and recognition of every person regardless of their social status, hierarchical position, national, religious, or sexual affiliation. Based on the research of Xu et al. [2], they see the opportunity in algorithmic recruiting in that the discriminatory tendencies and unfairness can be prevented, as correctly programmed algorithms are directly linked to consistent and unbiased decisions.

Some authors, such as Köchling et al. [48] proposed ways for measuring unfairness or establishing fairness as a post- or pre-process procedure to compare the usability in a recruiting context. Hajian et al. [49] made very technical proposals to guarantee fair algorithms, data mining especially. They mentioned some pre-processing approaches, such as logistic regression and naive bayes models, as well as post-processing approaches including for example simultaneous discrimination prevention and privacy protection considering k-anonymity or rule and pattern mining. As underlined by Kuleshov & Abramova [50] there is a need to avoid and prevent any forms of discrimination and inequality associated with the use of AI-based systems. This also applies to discrimination using algorithms to hire employees.

Discrimination by proxy means that a company or an institute discriminates in favor or against minorities [2]. Raghavan et al. [6] explain that while recruiting

decisions of algorithms have predictive validity, they often disadvantage minorities, even if the candidate has the same potential as their white counterpart. Hajian et al. [49] describe that even if there is no discriminatory intent, it is possible that big data can have discriminatory effects further imposing the less favorable treatment to already disadvantaged groups. The authors present two approaches to a solution: 1) techniques for discrimination discovery from databases and 2) discrimination prevention by means of fairness-aware data mining by developing data mining systems that are discrimination-conscious by-design.

As already explained, discrimination include actions directed against minorities in the vast majority of cases. This includes racial and ethnic groups as well as diverse groups. Racial discrimination appears as racial bias in algorithms that are trained on biased data.

Xu and colleagues [2] combine unfairness and racial discrimination by explaining that especially video interviewing includes race as the most controversial characteristics. That is because algorithms are, and have been, mostly trained with data of white people, e.g., images of white people which results in negative consequences for people of color. Köchling & Wehner [15] concluded that all races that were disadvantaged in history, e.g., Hispanics, will have higher chances to be discriminated against by AI-based algorithms trained with historical data. The reason for biases in algorithms is related to the input data is that certain groups or characteristics are mostly underrepresented or sometimes overrepresented. This phenomenon is also called representation bias. Likewise, Kuleshov & Abramova [50] and Ghallab [36] found that the racial diversity of humanity, as well as associated differences in social status and ethnicity influenced the training data of the algorithms in such a way that the results were affected accordingly beforehand.

Diversity and inclusion, related to e.g., disabled people or members of the LGBTQ+ community, is increasingly lived today and, according to Raghavan et al. [6] and Ghallab [35], should not be missing in the design of ethical guidelines. There is the importance of respecting the diversity of humanity and preventing discrimination of any kind.

4.2 Ethics guidelines for AI

After categorizing the articles into ethical principles for AI-based recruiting algorithms, we considered which existing guidelines for AI already included aspects of these principles. Within our basket of 26 articles, we identified 15 types of ethics guidelines for AI. The ethics guidelines for AI

including the principle of fairness are summarized in Table 2.

Table 2. Ethics guidelines including the principle of fairness

Author(s)	Existing guidelines
AI Now, 2016	The AI Now Report
Cutler et al., 2019	Everyday Ethics for Artificial Intelligence
Diakopoulos et al., 2019	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms
Google, 2019	Perspectives on Issues in AI Governance
Kuleshov & Abramova, 2020	Addressing AI ethics through codification
Microsoft, 2021	Responsible AI principles from Microsoft
OpenAI, 2018	OpenAI Charter
Pekka et al., 2018	EU Ethics guidelines for trustworthy AI

Most of the already existing guidelines have dealt with the aspect of fairness. It is an overarching and open concept, which is why eight of the 15 guidelines clearly addressed it.

Pekka et al. [11] said in their EU-Guidelines that the “regulation of AI systems must be fair” (p.10) and that this is one of the most important principles. They refer to the developers of AI systems, who are supposed to make sure that minorities are considered during the implementation. They include bias, stigmatisation, and discrimination as important aspects of unfairness. Other guidelines also imply race, ethnicity, gender, nationality as well as income, sexual orientation, ability, and political or religious belief as factors which evoke unfairness [51,52]. Most importantly, it is essential to be conscious of the danger that AI algorithms and the underlying data can not only reflect but also reinforce unfairness. However, it is also possible to reduce prejudices and unfair bias [52].

Diakopoulos et al. [53] provided concrete recommendations on how the aspect fairness can be considered. They suggest establishing a certain level of fairness awareness to enable systems to correct errors that are related to different categorization for example of populations. Furthermore, the false positive and false negative assignments need to be calculated. The guidelines by OpenAI [54] stressed

that instead of replicating errors, it is important to first detect and then correct them. Moreover, another way to approach the problem is to create software that explains the errors that lead to unfairness so that they can be understood and corrected.

Cutler et al. [55] believe that the development team has a responsibility to minimize unfairness, because they, like every other human, are inherently vulnerable to biases and could implement them to the algorithm.

To avoid unfairness AI Now [56] and Kuleshov & Abramova [50] propose to focus on the data. It is of great importance to use representative datasets. More precisely, this means that the data has to be unbiased and from reliable and legal sources. With this data the algorithm has to be tested several times and there needs to be a control that no unfair bias appears.

Furthermore, we summarized the ethics guidelines which are related to the aspect of discrimination in Table 3.

Table 3. Ethics guidelines including the principle of discrimination

Author(s)	Existing guidelines
Accenture, 2021	An Ethical Framework for Responsible AI and Robotics
Amershi et al., 2019	Guidelines for Human-AI Interaction
Beijing Academy of Artificial Intelligence, 2020	Beijing Academy of Artificial Intelligence
Kuleshov & Abramova, 2020	Addressing AI ethics through codification
Organisation for Economic Co-operation and Development, 2019	OECD Legal Instruments Recommendation of the Council on Artificial Intelligence
Pekka et al., 2018	EU Ethics guidelines for trustworthy AI

The avoidance of discrimination was also mentioned in eight out of 15 ethic guidelines and is therefore a very important principle. Discrimination is often mentioned as subitem of fairness. For this reason, it is not always intensively executed in the guidelines but enumerated as part that should always be considered. The guidelines outline that its relevant to try the best to reduce discrimination in AI [57-59].

Abrassart et al. [60] establish two principles dealing with discrimination. First, the algorithm

should never support bias, prejudices nor inequity. From the start of the developing process to the training with data at the end, discrimination has to be considered as critical influence. Second, not only the creation, reinforcement and reproduction are important while designing the algorithm, but also the elimination of all kinds of discrimination. Another aspect of discrimination is the usage of language, that should never exclude or discriminate [61].

The EU guidelines for Trustworthy AI categorized equality as a stronger peculiarity of discrimination, “which tolerates the drawing of distinctions between dissimilar situations based on objective justifications. In an AI context, equality [...] also requires adequate respect of inclusion of minorities, traditionally excluded” [11, p.7]. Discrimination in AI could be either intentionally or unintentionally, but both has to disappear as soon as possible, because it disadvantages certain groups.

We also found ethics guidelines which discussed more on the avoidance of bias in AI-based systems. We summarized them in Table 4.

Table 4. Ethics guidelines including the principle of the avoidance of bias

Author	Name of guideline
AI Now, 2016	The AI Now Report
Amershi et al., 2019	Guidelines for Human-AI Interaction
Beijing Academy of Artificial Intelligence, 2020	Beijing Academy of Artificial Intelligence
Cutler et al., 2019	Everyday Ethics for Artificial Intelligence
Diakopoulos et al., 2019	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms
Kuleshov & Abramova, 2020	Addressing AI ethics through codification
Pekka et al., 2018	EU Ethics guidelines for trustworthy AI
Sage, 2020	The Ethics of Code

As a part of discrimination, race, diversity and sexual orientation were mentioned and explained more specifically in nine out of 15 Guidelines.

In terms of race and the racial bias that can occur, Cutler et al. [55] emphasized in their IBM guidelines that it could be beneficial to involve policymakers or academics, that are familiar with the topic and problem

of racial bias and are able to include a new perspective. Moreover, it is always important to realize the own culture and related value system in comparison to others, to integrate this difference in the design process. Diakopoulos et al. [53] developed a guiding question and corresponding initial steps to take.

Overall, diversity of humans needs to always be considered and gain more attention in the process of designing an algorithm, including validation and testing of the algorithm [e.g., 11]. Besides this, the team is also an important factor and should be diverse as mentioned in the ethics guidelines by Sage [62] and AI Now [63].

Moreover, Abrassart et al. [60] describe, that it is necessary to “[...] avoid using acquired data to lock individuals into a user profile, fix their personal identity, or confine them to a filtering bubble, which would restrict and confine their possibilities for personal development” (p.14). It is important to address both genders [him, her] without attributing certain characteristics to either of them [60,61]. There are several diverse sexual, cultural, or social expressions that have to be taken into account when designing AI.

To ensure optimal coverage of the three categories of ethical principles, we need to take parts from all existing in order to derive guidance for using AI-based algorithms in recruiting.

5. Discussion and Guidance for Information Systems Research

One of the most fundamental principles to avoid an unfair and biased recruiting algorithm are the trainings data, that are used. Several guidelines that were analyzed in the literature review mention the crucial role of the data collected and used for the training of AI algorithms [11, 50, 53, 54, 57, 58, 60]. It is important to reduce the unintentional bias in the data as far as possible so that the algorithm does not discriminate. That this is not an easy task and can be a big challenge. Nevertheless, it is essential to collect data in a way that avoids unfairness and discrimination, because AI algorithms identify patterns from trainings data and will therefore adopt bias. Therefore, it is important to identify possible bias as early as the data selection stage [51]. We derived the following guideline:

- (1) *Organizations should use sample data sets as training data, that are adjusted and balanced by human diversity experts. Those experts should randomly manually check training datasets for their level of diversity.*

Therefore, it is necessary to take a closer look at the demographics of the company, because often data is used that is based on historical data of the companies. But if a company, for example, has hired mostly white males in the past, this will be reflected in the data and the algorithm will adjust accordingly. Thus, it is important that the algorithms are not trained on historical data of the company but rather with data that is repeatedly validated. The EU Guidelines recommend using an “[...] AI technology that can be employed to identify this inherent bias, and hence to support awareness training on our own inherent bias” [11] p.16. Also, Diakopoulos et al. [53] suggest a fairness-aware data mining algorithm to try to reduce patterns that create, reinforce, or reproduce unfairness. Therefore, we recommend following guideline:

- (2) *Because machine learning systems can entrench existing bias in decision making systems, AI-based recruitment algorithms need to be designed, researched and trained in such a way that they that they make discrimination reproducible checking for — among other things — social, sexual, ethnic, cultural, or religious differences. Organizations need to deploy fairness-aware data mining algorithms that can identify the inherent bias.*

To ensure that all stakeholders are represented in the development, implementation and evaluation of an AI application, the respective responsible teams need to mirror the environment which the application will work in. As a result, multiple points of views will be included and preventative measures to ensure an unbiased implication can be achieved. There is a great inequality when looking at the male-female ratio in AI research. 80% of AI professors in prestigious US universities are men [63], only 18% of authors at the 21 leading AI conferences are women, 71% of applicants for AI roles in the US in 2017 were men and only 22% of AI professionals on LinkedIn are women with no evidence of improvement in recent years. This gender gap might be the most obvious but seeing that there is little improvement, other historically disadvantaged groups are even less involved in the AI implementation process. For this reason, guidelines that demand to “increase efforts to improve diversity among AI developers and researchers [...]” [56] are not sufficient anymore. Instead, we propose to implement following guidelines:

- (3) *Organizations should ensure that those people who design an AI-based recruitment system, whether in research or industry, reflect*

inclusiveness and diversity of the individuals and groups of the society. Unless we build AI using diverse teams, data sets and design, we are at risk of repeating the inequality of previous revolutions.

Starting points where researchers could intervene could be shedding light on the culture e.g., multiculturalism, that establishes the algorithms and raise awareness and articulate relevant perspectives. Another aspect is the support of research to develop the means of measuring and assessing AI systems' accuracy and fairness during the design and deployment stage. A third starting point could be a feedback mechanism or an open dialogue system which takes user-identified biases into account. This could be realized through feedback loops e.g., chatbots, who ask for an opinion after the algorithm suggested something. As mentioned above, to guarantee independency and have an objective part in the development experts with great knowledge are crucial. Therefore, we recommend the following guideline:

- (4) *Ethics experts, external professionals, or persons with a similar position need to be included in the whole recruitment process and moreover need to be capable of measuring errors and (unintended) biases in algorithms at any time. This can ensure the AI-based recruiting systems' accuracy and fairness is in accordance to ethical standards of societies. They not only need to detect the errors and biases but be able to mitigate them and support the company by preventing further mistakes when applying AI-based recruitment algorithms.*

6. Conclusion

Discrimination in algorithms is a ubiquitous issue in the recruitment process of companies. The goal of this review was to identify which criteria are essential in recruiting algorithms. Subsequently, which existing ethics guidelines take these criteria into account.

Thus, as the contribution of this paper we identified principles that should not be missing in future ethics guidelines for recruiting algorithms. We identified three categories of relevant principles: 1) Fairness 2) the prevention of discrimination and 3) the avoidance of bias. We also identified three sub-aspects for the third principle: 1) racial bias 2) diversity, and 3) sexual orientation. Furthermore, we highlighted which existing ethics guidelines on AI covered some of these aspects. With this review, the amount of ethics guidelines was reduced to the most important criteria

in relation to recruiting algorithms, thus a good overview could be achieved. Based on these guidelines, we derived more concrete guidelines for deploying AI-based recruiting algorithms in a company.

However, there are limitations to this research. The question of the impact of ethics guidelines remains a major point of discussion, as they are merely recommendations and orientations and are not enshrined in law [45]. There is no knowledge or evidence, that ethics guidelines do have impact in the industry [2]. Further research could therefore investigate to what extent ethics guidelines have an impact on the industry and how they are best formulated to guarantee efficient implementation.

There are already various tools for implementing AI ethics principles, but they are not sufficient or only partially specific [2]. Therefore, in further research it is useful to focus on methods and implementation strategies for AI ethics based on the ethics principles outlined in this paper. However, the specification of ethics guidelines for recruiting algorithms as industry-specific codes of practice would facilitate classification, evaluation and measurement of systems as well as reduce and prevent discrimination and biases, both at the technical level and at the level of public perception and trust [50]. Even if the ethics guidelines impact the industry, it is necessary to keep in mind, that algorithms that use AI are mostly so complex that it is impossible to check parameters, computational logic and analyze input data in terms of biases.

Considering the methodological aspects of this paper, we could not completely ensure that important aspects might have been overlooked. The search was limited English keywords, resulting in articles published in other languages were not included. Even if we chose a keyword and backward search as well as an iterative process, there is a risk of overlooking relevant literature, because similar phenomena are often studied using several synonyms.

In summary, with our SLR, the extraction of relevant ethics principles for AI-based recruiting algorithms and with four initial guidelines for organizations that plan to use recruiting algorithms, we would like to encourage scholars to further elaborate on this field as the relevance of AI-based recruiting algorithms will increase in the near future.

References

- [1] Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, 63(2), 215–226.

- [2] Xu, P., Barbosa, D., Riley, P., McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., Ahmed, S. M., Kuleshov, A., Ignatiev, A., Abramova, A., Marshalko, G., Köchling, A., Wehner, M. C., Riazzy, S., Wehner, M. C., Simbeck, K., Ghallab, M., ... Abrassart, C. (2020). AI and recruiting software: Ethical and legal implications. *AI Perspectives*, 43(1), 795–848.
- [3] Qu, L. (2019). 99% of Fortune 500 Companies use Applicant Tracking Systems (ATS) - *Jobscan Blog*. <https://www.jobscan.co/blog/99-percent-fortune-500-ats/> (Last accessed: 2021/08/30).
- [4] Lang, S., Laumer, S., Maier, C., & Eckhardt, A. (2011). Drivers, challenges and consequences of E-recruiting: A literature review. *SIGMIS CPR 2011 - Proceedings of the 2011 ACM SIGMIS Computer Personnel Research Conference*, 26–35.
- [5] Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to “solve” the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 458–468.
- [6] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481.
- [7] Bendick, M., & Nunes, A. P. (2012). Developing the Research Basis for Controlling Bias in Hiring. *Journal of Social Issues*, 68(2), 238–262.
- [8] Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4), 991–1013.
- [9] Johnson, S. K., Hekman, D. R., & Chan, E. (2016). If there’s only one woman in your candidate pool, there’s statistically no chance she’ll be hired. *Harvard Business Review*, 26(4).
- [10] Romero, J. H., Bendick, M., & Jackson, C. W. (1997). Employment discrimination against older workers: An experimental study of hiring practices. *Journal of Aging and Social Policy*, 8(4), 25–46.
- [11] Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and A path forward. *California Management Review*, 61(4).
- [12] Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.
- [13] Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848.
- [14] Laurim, V., S. Arpacı, B. Prommegger, and H. Krcmar, “Computer, whom should I hire? - Acceptance criteria for artificial intelligence in the recruitment process”, *Proceedings of the Annual Hawaii International Conference on System Sciences 2020-January*, 2021, pp. 5495–5504.
- [15] Webster, J., and T.R. Watson, “Analyzing the Past to Prepare for the Future: Writing a Literature Review”, *MIS Quarterly* 26(2), 2002.
- [16] Li, L., P. Gao, and J.Y. Mao, “Research on IT in China: A call for greater contextualization”, *Journal of Information Technology* 29(3), 2014, pp. 208–222.
- [17] Möhlmann, M., & Zalmanson, L. (2018). Hands on the Wheel: Navigating Algorithmic Management and Uber Drivers Autonomy. *ICIS 2017: Transforming Society with Digital Innovation*.
- [18] Daugherty, P. R., & Wilson, H. J. (2018). Human + Machine: Reimagining Work in the Age of AI. In *Hilos Tensados*.
- [19] Savage, D., & Bales, R. (2017). Video games in job interviews: using algorithms to minimize discrimination and unconscious bias. *ABA Journal of Labor & Employment Law*, 32(2).
- [20] Suen, H. Y., Chen, M. Y. C., & Lu, S. H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, 98.
- [21] Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1).
- [22] Langer, M., König, C. J., & Papanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3).
- [23] Akerlof, G. A., & Shiller, R. J. (2010). Animal spirits: How human psychology drives the economy, and why it matters for global capitalism. In *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*.
- [24] Barocas, S., & Selbst, A. (2016). Big Data’s Disparate Impact. *California Law Review*, 104(3).
- [25] Sweeney, L. (2013). Discrimination in online Ad delivery. *Communications of the ACM*, 56(5).
- [26] Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*.
- [27] Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, 5(1).
- [28] Kim, P. T. (2017). Data-Driven Discrimination at Work. In *William & Mary Law Review* (Vol. 58).
- [29] Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning.
- [30] Koolen, C., & van Cranenburgh, A. (2017). *These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution*.
- [31] Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*.
- [32] Arrow, K. J. (2015). The theory of discrimination. In *Discrimination in Labor Markets*.

- [35] Ghallab, M. (2019). Responsible AI: requirements and challenges. *AI Perspectives*, 1(1).
- [36] Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race and recidivism: predictive bias and disparate impact. *Criminology*, 54(4).
- [37] Cappelli, P., & Tavis, A. (2016). The performance management revolution. In *Harvard Business Review* (2016, October).
- [38] DeNisi, A., & Smith, C. E. (2014). Performance Appraisal, Performance Management, and Firm-Level Performance: A Review, a Proposed Model, and New Directions for Future Research. *Academy of Management Annals*, 8(1).
- [39] Schoorman, F. D. (1988). Escalation bias in performance appraisals: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology*, 73(1).
- [40] Bryson, J., & Winfield, A. (2017). Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. *Computer*, 50(5).
- [41] Etzioni, A., & Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. *Journal of Ethics*, 21(4).
- [42] Someh, I. A., Breidbach, C. F., Davern, M., & Shanks, G. (2016). Ethical implications of big data analytics. *24th European Conference on Information Systems, ECIS 2016*.
- [43] Hagedorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1).
- [44] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- [45] Rothenberger, L., Fabian, B., & Arunov, E. (2019). *Relevance of ethical guidelines for artificial intelligence - a survey and evaluation*. Progress Papers.
- [46] vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, 37, 205–224.
- [47] Ochmann, J., & Laumer, S. (2019). Fairness as a Determinant of AI Adoption in Recruiting: An Interview-based Study. *DIGIT 2019 Proceedings*.
- [48] Köchling, A., Riaz, S., Wehner, M. C., & Simbeck, K. (2021). Highly Accurate, But Still Discriminatory: A Fairness Evaluation of Algorithmic Video Analysis in the Recruitment Context. *Business and Information Systems Engineering*, 63(1), 39–54.
- [49] Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Aug, 2125–2126*.
- [50] Kuleshov, A., & Abramova, A. (2020). Addressing AI ethics through codification. *International Conference Engineering Technologies and Computer Science (EnT)*.
- [51] Google. (2019). *Perspectives on Issues in AI Governance*. <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf> (Last accessed: 2021/08/30).
- [52] Microsoft. (2021). *Responsible AI principles from Microsoft*. <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6>
- [53] Diakopoulos, N., Friedler, S. A., Arenas, M., Barocas, S., Hay, M., Howe, B., & Jagadish, H. V. (2019). *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms: FAT ML*.
- [54] OpenAI. (2018). *OpenAI Charter*. <https://openai.com/charter/> (Last accessed: 2021/08/30).
- [55] Cutler, A., Pribić, M., & Humphrey, L. (2019). *Everyday Ethics for Artificial Intelligence*.
- [56] AI Now. (2016). *The AI Now Report the Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term AI Now Overview*.
- [57] Accenture. (2021). *An Ethical Framework for Responsible AI and Robotics*. <https://www.accenture.com/gb-en/company-responsible-ai-robotics>
- [58] Beijing Academy of Artificial Intelligence. (2020). *Beijing AI Principles*. <https://www.baai.ac.cn/news/beijing-ai-principles-en.html> (Last accessed: 2021/08/30).
- [59] Organisation for Economic Co-operation and Development. (2019). *OECD Legal Instruments Recommendation of the Council on Artificial Intelligence*.
- [60] Abrassart, C., Bengio, Y., Chicoisne, G., de Marcellis Warin, N., Dilhac, M.-A., Gambis, S., & Gautrais, V. (2018). *Montréal Declaration for a responsible development of artificial intelligence*.
- [61] Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz Microsoft, E., Allen, P. G., & Fournay, A. (2019). *Guidelines for Human-AI Interaction*. 13.
- [62] Sage Group, *The Ethics of Code: Core Principles*, 2017. <https://www.sage.com/~media/group/files/business-builders/business-builders-ethics-of-code.pdf> (Last accessed: 2021/08/30).
- [63] Lecher, C., “The artificial intelligence field is too white and too male, researchers say”, *The Verge*, 2019. <https://www.theverge.com/2019/4/16/18410501/artificial-intelligence-ai-diversity-report-facial-recognition#:~:text=Women and people of color,percent at Google are women.> (Last accessed: 2021/08/30).