



REVIEW ARTICLE



<https://doi.org/10.1057/s41599-023-02079-x>

OPEN

# Ethics and discrimination in artificial intelligence-enabled recruitment practices

Zhisheng Chen <sup>1</sup>✉

This study aims to address the research gap on algorithmic discrimination caused by AI-enabled recruitment and explore technical and managerial solutions. The primary research approach used is a literature review. The findings suggest that AI-enabled recruitment has the potential to enhance recruitment quality, increase efficiency, and reduce transactional work. However, algorithmic bias results in discriminatory hiring practices based on gender, race, color, and personality traits. The study indicates that algorithmic bias stems from limited raw data sets and biased algorithm designers. To mitigate this issue, it is recommended to implement technical measures, such as unbiased dataset frameworks and improved algorithmic transparency, as well as management measures like internal corporate ethical governance and external oversight. Employing Grounded Theory, the study conducted survey analysis to collect firsthand data on respondents' experiences and perceptions of AI-driven recruitment applications and discrimination.

## Introduction

Technological innovation has revolutionized work across the first through fourth industrial revolutions. The fourth industrial revolution introduced disruptive technologies like big data and artificial intelligence (Zhang and Chen, 2023). The advancement of data processing and big data analytics, along with developments in artificial intelligence, has improved information processing capabilities, including problem-solving and decision-making (Raveendra et al., 2020). With the increasing normalization and timely usage of digital technologies, there is a potential for future higher-level implementation of AI systems (Beneduce, 2020).

AI can provide faster and more extensive data analysis than humans, achieving remarkable accuracy and establishing itself as a reliable tool (Chen, 2022). It can collect and evaluate large amounts of data that may exceed human analytical capacities, enabling AI to provide decision recommendations (Shaw, 2019).

Modern technologies, including artificial intelligence solutions, have revolutionized work and contributed to developing human resources management (HRM) for improved outcomes (Hmoud and Laszlo, 2019). One significant area where their impact is felt is in the recruitment process, where AI implementation can potentially provide a competitive advantage by enabling a better understanding of talent compared to competitors, thereby enhancing the company's competitiveness (Johansson and Herranen, 2019).

AI receives commands and data input through algorithms. While AI developers believe their algorithmic procedures simplify hiring and mitigate bias, Miasato and Silva (2019) argue that

<sup>1</sup>College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China. ✉email: [njucz@nuaa.edu.cn](mailto:njucz@nuaa.edu.cn)

algorithms cannot eliminate discrimination alone. The decisions made by AI are shaped by the initial data it receives. If the underlying data is unfair, the resulting algorithms can perpetuate bias, incompleteness, or discrimination, creating potential for widespread inequality (Bornstein, 2018). Many professionals assert that AI and algorithms reinforce socioeconomic divisions and expose disparities. To quote Immanuel Kant, “In the bentwood of these data sets, none of them is straight” (Raub, 2018). This undermines the principle of social justice, causing moral and economic harm to those affected by discrimination and reducing overall economic efficiency, leading to decreased production of goods and services.

AI recruitment tools have a concerning aspect that cannot be overlooked, highlighting the need to address these challenges through technical or managerial means (Raub, 2018). Increasing evidence suggests that AI is more impartial than commonly believed; however, algorithms and AI can result in unfair employment opportunities and the potential for discrimination without accountability. To harness the benefits of AI in recruiting, organizations should exercise careful selection of their programs, promote the adoption of accountable algorithms, and advocate for improvements in racial and gender diversity within high-tech companies.

The general construct of this study is, first, an extension of statistical discrimination theory in the context of the algorithmic economy; second, a synthesis of the current literature on the benefits of algorithmic hiring, the roots and classification of algorithmic discrimination; and third, initiatives to eliminate the existence of algorithmic hiring discrimination; fourth, based on the Grounded Theory, we conduct surveys with respondents and analyze primary data to support the study.

The contributions of this study are as follows:

First, discuss job market discrimination theories in the digital age context. When considering statistical discrimination theories, we should consider the current circumstances. It is necessary to apply these discrimination theories to evaluate the issues that arise from the use of technology in the digital age, particularly with the widespread adoption of artificial intelligence, big data, and blockchain across various industries.

Secondly, a literature review approach was employed to examine the factors contributing to discrimination in algorithmic hiring. Our goal with this analysis is to help managers and researchers better understand the limitations of AI algorithms in the hiring process. We conducted a thorough review of 49 papers published between 2007 and 2023 and found that there is currently a fragmented understanding of discrimination in algorithmic hiring. Building on this literature review, our study aims to offer a comprehensive and systematic examination of the sources, categorization, and possible solutions for discriminatory practices in algorithmic recruitment.

Thirdly, we take a comprehensive approach that considers technical and managerial aspects to tackle discrimination in algorithmic hiring. This study contends that resolving algorithmic discrimination in recruitment requires technical solutions and the implementation of internal ethical governance and external regulations.

The subsequent study is structured into five parts. The first section provides the theoretical background for this research. The following section outlines the research methodology employed in the literature review and identifies four key themes. The third section delves into a detailed discussion of these four themes: applications and benefits of AI-based recruitment, factors contributing to algorithmic recruitment discrimination, types of discrimination in algorithmic recruitment, and measures to mitigate algorithmic hiring discrimination. The fourth section involves conducting a survey among respondents and analyzing

the primary data collected to support our study. The final section concludes by suggesting future directions for research.

### Theory background

**Discrimination theory.** Discrimination in the labor market is defined by the ILO’s Convention 111, which encompasses any unfavorable treatment based on race, ethnicity, color, and gender that undermines employment equality (Ruwanpura, 2008). Economist Samuelson (1952) offers a similar definition, indicating that discrimination involves differential treatment based on personal characteristics, such as ethnic origin, gender, skin color, and age.

Various perspectives on the causes and manifestations of discrimination can be broadly categorized into four theoretical groups. The first is the competitive market theory, which explains discriminatory practices within an equilibrium of perfect competition (Lundberg and Startz, 1983). This view attributes discrimination primarily to personal prejudice. The second is the monopoly model of discrimination, which posits that monopolistic power leads to discriminatory behavior (Cain, 1986). The third is the statistical theory of discrimination, which suggests that nonobjective variables, such as inadequate information, contribute to biased outcomes (Dickinson and Oaxaca, 2009). Lastly, we have the antecedent market discrimination hypothesis as the fourth category.

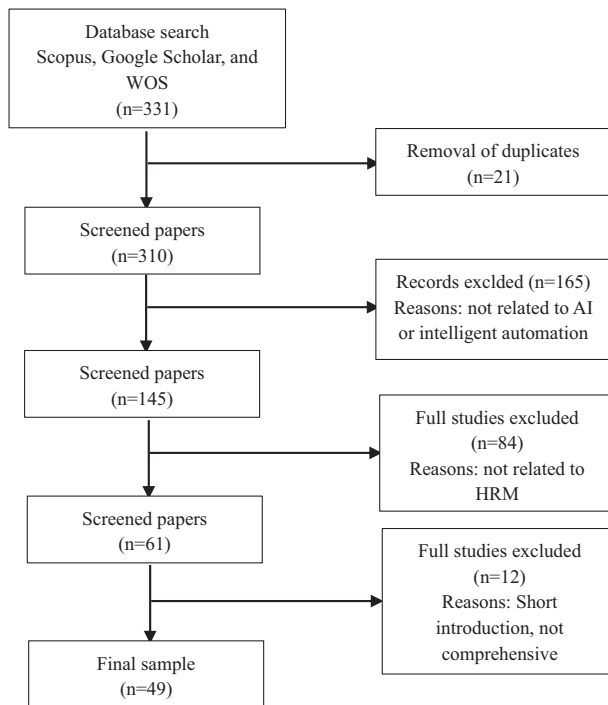
**Statistical discrimination theory.** Statistical discrimination refers to prejudice from assessment criteria that generalize group characteristics to individuals (Tilcsik, 2021). It arises due to limitations in employers’ research techniques or the cost constraint of obtaining information in the asymmetry between employers and job seekers. Even without monopolistic power, statistical discrimination can occur in the labor market due to information-gathering methods. Employers are primarily interested in assessing candidates’ competitiveness when making recruitment decisions. However, obtaining this information directly is challenging, so employers rely on various indirect techniques.

Discrimination carries both individual and societal economic costs. The social cost arises from the decrease in overall economic output caused by discrimination. However, this is still deemed efficient under imperfect information and aligns with the employer’s profit maximization goal. Therefore, it is likely that statistical discrimination in employment will persist.

### Extension of statistical discrimination theory in the digital age.

The digital economy has witnessed the application of various artificial intelligence technologies in the job market. Consequently, the issue of algorithmic hiring discrimination has emerged, shifting the focus of statistical discrimination theory from traditional hiring to intelligent hiring. The mechanisms that give rise to hiring discrimination problems remain similar, as both rely on historical data of specific populations to predict future hiring outcomes.

While AI recruiting offers numerous benefits, it is also susceptible to algorithmic bias. Algorithmic bias refers to the systematic and replicable errors in computer systems that lead to unequal and discrimination based on legally protected characteristics, such as race and gender (Jackson, 2021). When assessments consistently overestimate or underestimate a particular group’s scores, they produce “predictive bias” (Raghavan et al., 2020). Unfortunately, these discriminatory results are often overlooked or disregarded due to the misconception that AI processes are inherently “objective” and “neutral.”



**Fig. 1 Screening flow chart.** Procedures used in the literature review to reveal emerging themes.

Despite algorithms aiming for objectivity and clarity in their procedures, they can become biased when they receive partial input data from humans. Modern algorithms may appear neutral but can disproportionately harm protected class members, posing the risk of “agentic discrimination” (Prince and Schwarcz, 2019). If mishandled, algorithms can exacerbate inequalities and perpetuate discrimination against minority groups (Lloyd, 2018).

Within the recruitment process, algorithmic bias can manifest concerning gender, race, color, and personality.

### Research methodology

The primary research strategy was a literature review approach. This review aimed to assess current research on recruitment supported by artificial intelligence algorithms. The systematic review process included gathering and evaluating the selected studies’ literature and topics. Driven by the direction of the research, studies focusing on algorithmic discrimination in recruitment over the past 10 years were included unless past literature was worth reviewing. This is because this is a relatively new phenomenon that has become prominent over the past 10 years. In defining the “algorithmic and hiring discrimination” literature, a fairly broad approach was taken based on article keywords rather than publication sources. Depending on the focus, keywords related to algorithms and hiring discrimination were included in the search string. The keyword search algorithm for this review is as follows. (“artificial intelligence” and “hiring discrimination”), (“algorithms” and “recruitment discrimination”), (artificial intelligence” and “recruitment discrimination”), and (“algorithms” and “hiring discrimination”). SCOPUS, Google Scholar, and Web of Science are three well-known search engines frequently used by the academic community and meet the criteria for technology-related topics in this review. WOS is used as a starting point for high-quality peer-reviewed scholarly articles. The study selected these three databases, used search engines, and maintained ten years. After applying an initial screening related to titles, keywords, or abstracts, the literature was selected based on its relevance to the research topic.

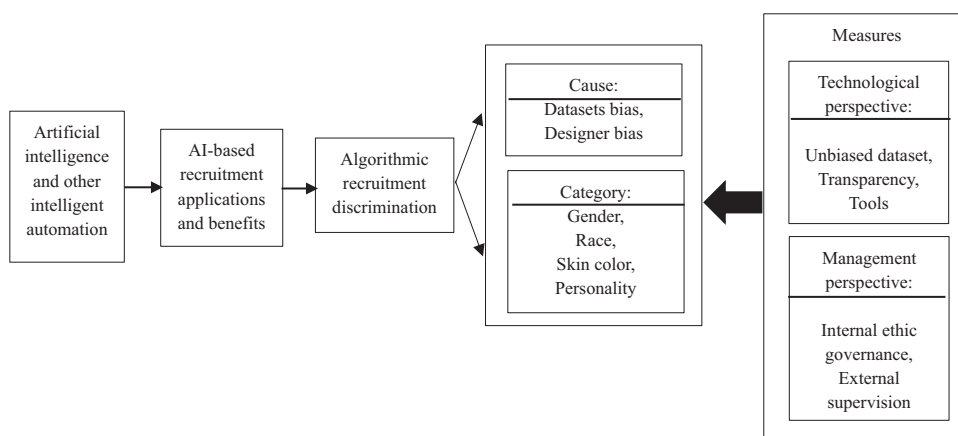
The obtained literature was studied in depth to reveal the surfaced themes. Several systematic research themes were identified, including AI-based recruitment applications and benefits, causes of algorithmic discrimination, which algorithmic recruitment discrimination exists, and algorithmic recruitment discrimination resolution.

The process applied for the reviews depicted in Fig. 1. After excluding duplicates and less relevant and outdated literature, only 45 articles could be used as references for this study (referred to Table 1). The literature review shows that most of the research on algorithmic hiring discrimination has occurred in recent years. The research trend indicates that algorithmic hiring discrimination will be a hot research topic in the coming period.

1. The first theme is the application of various aspects of recruitment based on artificial intelligence support and its benefits. Bogen and Rieke (2018), Ahmed (2018), Hmoud and Laszlo (2019), Albert (2019), van Esch et al. (2019), Köchling et al. (2022), and Chen (2023) consider the recruitment process as a set of tasks that may be divided into four steps sourcing, screening, interviewing, and selection. Each step includes different activities, and AI algorithms can change how each stage is executed. Some studies point out that AI-supported recruitment has benefits. Beattie et al. (2012), Newell (2015), Raub (2018), Miasato and Silva (2019), Beneduce (2020), and Johnson et al. (2020) state that it can reduce costs; Hmoud and Laszlo (2019), Johansson and Herranen (2019), Raveendra et al. (2020), Black and van Esch (2020), and Allal-Chérif et al. (2021) suggest it saves time; Upadhyay and Khandelwal (2018) and Johansson and Herranen (2019) present it reducing transactional workload.
2. The second theme is the causes of algorithmic discrimination. McFarland and McFarland (2015), Mayson (2018), Raso et al. (2018), Raub (2018), Raghavan et al. (2020), Njoto (2020), Zixun (2020), and Jackson (2021) suggest that the reason for algorithmic discrimination is related to data selection. Data collection tends to prefer accessible, “mainstream” organizations unequally dispersed by race and gender. Inadequate data will screen out groups that have been historically underrepresented in the recruitment process. Predicting future hiring outcomes by observing historical data can amplify future hiring inequalities. Yarger et al. (2019), Miasato and Silva (2019), and Njoto (2020) propose that discrimination is due to the designer-induced selection of data features.
3. The third theme is which algorithmic recruitment discrimination exists. According to Correll et al. (2007), Kay et al. (2015), O’neil (2016), Raso et al. (2018), Miasato and Silva (2019), Langenkamp et al. (2019), Faragher (2019), Ong (2019), Fernández and Fernández (2019), Beneduce (2020), Jackson (2021), Yarger et al. (2023), and Avery et al. (2023), when partial human data is provided to a machine, so the algorithm is biased, it will eventually lead to the risk of “agent discrimination.” In recruitment, algorithmic bias can manifest in gender, race, skin color, and personality.
4. The fourth theme is algorithmic recruitment discrimination resolution. Kitchin and Laurialt (2015), Bornstein (2018), Raso et al. (2018), Xie et al. (2018), Raub (2018), Bornstein (2018), Grabovskyi and Martynovych (2019), Amini et al. (2019), Shin and Park (2019), Yarger et al. (2019), Gulzar et al. (2019), Kessing (2021) Jackson (2021), and Mishra (2022) argue that fair data sets need to be constructed and algorithmic transparency needs to be improved. Moreover, Smith and Shum (2018), Mitchell et al. (2019), Ong (2019), Zuiderveen Borgesius (2020), Peña et al. (2020), Kim et al.

**Table 1 Research theme, papers, and main content.**

Theme	Research papers	Main content
1. AI-enabled recruitment application and its benefits.	Bogen and Rieke (2018), Ahmed (2018), Hmoud and Laszlo (2019), Albert (2019), van Esch et al. (2019), Köchling et al. (2022), and Chen (2023) Beattie et al. (2012), Newell (2015), Raub (2018), Miasato and Silva (2019), and Beneduce (2020), Johnson et al. (2020). Hmoud and Laszlo (2019), Johansson and Herranen (2019), Raveendra et al. (2020), Black and van Esch (2020), and Allal-Chérif et al. (2021) Upadhyay and Khandelwal (2018), Johansson and Herranen (2019)	Summarize the recruitment process into four main steps: sourcing, screening, interviewing, and selection Claim that AI-supported recruitment can reduce costs Suggests that AI support can save time Propose that AI support can reduce transactional workload
2. Causes of algorithmic discrimination	McFarland and McFarland (2015), Mayson (2018), Raso et al. (2018), Raub (2018), Raghavan et al. (2020), Njoto (2020), Zixun (2020), and Jackson (2021) Yarger et al. (2019), Miasato and Silva (2019), and Njoto (2020)	Suggests that the reason for algorithmic discrimination is related to data selection Propose that discrimination is due to the designer-induced selection of data features
3. Algorithmic recruitment discrimination	Correll et al. (2007), Kay et al. (2015), O’neil (2016), Raso et al. (2018), Miasato and Silva (2019), Langenkamp et al. (2019), Faragher (2019), Ong (2019), Fernández and Fernández (2019), Beneduce (2020), Jackson (2021), Yarger et al. (2023), and Avery et al. (2023)	In recruitment, algorithmic bias can manifest in gender, race, skin color, and personality
4. Algorithmic recruitment discrimination resolution	Kitchin and Laurialt (2015), Bornstein (2018), Raso et al. (2018), Xie et al. (2018), Raub (2018), Bornstein (2018), Grabovskiy and Martynovych (2019), Amini et al. (2019), Shin and Park (2019), Yarger et al. (2019), Gulzar et al. (2019), Kessing (2021), Jackson (2021), and Mishra (2022) Smith and Shum (2018), Mitchell et al. (2019), Ong (2019), Zuiderveen Borgesius (2020), Peña et al. (2020), Kim et al. (2021), Yang et al. (2021), and Jackson (2021)	Fair data sets and algorithmic transparency Propose internal ethical governance and external ethical oversight



**Fig. 2 An integrated conceptual framework summarizing the reviewed literature.** An overarching conceptual framework to visualize how AI and AI-based technologies can impact recruitment efforts.

(2021), Yang et al. (2021), and Jackson (2021) propose that from a management perspective, data governance need be strengthened, including internal ethical governance and external ethical oversight.

Through this review, we have created an overarching conceptual framework to visualize how AI and AI-based technologies impact recruitment. This framework is illustrated in Fig. 2 and aligns with the four research themes identified.

**Theme I. AI-based recruitment applications and benefits**

*Artificial Intelligence and algorithm.* The idea of machine intelligence dates back to 1950, when Turing, the father of computer science, asked, “Can machines think?” (Ong, 2019). The term “artificial intelligence” was coined by John McCarthy. Although

early scientists made outstanding contributions, artificial intelligence became an industry after the 1980s with hardware development. Initial applications of artificial intelligence were seen in the automation of repeated and complicated work assignments, like industrial robot production, that displaced human work in some plants. After the mid-1990s, artificial intelligence software saw significant advances. Until today’s digital economy, AI has been commonly used in various industries (Hmoud and Laszlo, 2019).

Artificial intelligence is defined as the ability of something like a machine to understand, learn, and interpret on its own in a human-like manner (Johansson and Herranen, 2019). Artificial intelligence aims “to understand and simulate human thought processes and to design machines that mimic this behavior.” It is designed to be a thinking machine with a level of human

intelligence (Jackson, 2021). However, large amounts of data must be combined with fast and iterative intelligent algorithms to handle this process, allowing ML systems to learn from patterns or features in the data automatically.

A set of instructions or commands used to carry out a particular operation is known as an algorithm. This digital process makes decisions automatically depending on the data entered into the program (Miasato and Silva, 2019). The algorithm analyzes massive data patterns through data mining, searching, and using ways to predict, like our point of view encoded in the code. It explores the dataset using agents representing various traits, such as race, sexual orientation, and political opinions (Njoto, 2020). The algorithms frequently contain these biases due to the lengthy history of racial and gender prejudices, both intentional and unconscious. When biases exist in algorithmic data, AI may replicate these prejudices in its decision-making, a mistake known as algorithmic bias (Jackson, 2021).

*AI-based recruitment stage.* The recruitment procedure is a series of events that may be divided into four significant steps: searching, screening, interviewing, and selection (Bogen and Rieke, 2018). Each phase includes various activities, and artificial intelligence technology can influence the execution of each stage.

The searching phase aims at a system for searching web content. It screens passive job applicants online through social media and recruitment platforms, analyzing their profiles according to predefined job descriptions. The search engine recognizes the meaning of the searched content and performs a web-based search to match candidates' profiles based on semantic annotations of job postings and profiles (Hmoud and Laszlo, 2019).

The screening phase includes evaluating and assessing the qualifications of candidates, where AI technology assists recruiters in scoring candidates and evaluating their competencies (Bogen and Rieke, 2018). The resumes are screened to match the job description better. The system can rank candidates according to the relevance of the qualification metrics.

The following phase is the interview. It is probably the most individual stage of the selection process and, thus, unlikely to be fully automated by artificial intelligence. However, some AI tools enable recruiters to conduct video interviews and research candidates' reactions, voice tones, and facial expressions (Ahmed, 2018).

The final stage is the selection stage, which is the stage where the employer makes the final employment decision. In this stage, AI systems can calculate remuneration and benefits for companies and anticipate the risk that candidates would violate workplace rules. (Bogen and Rieke, 2018).

### AI-based recruitment benefits

*Recruitment quality.* Beattie et al. (2012) found that some large companies believe unconscious bias affects recruitment quality. Organizations may need to hire more qualified people to avoid financial losses (Newell, 2015). Artificial intelligence has become a part of the recruitment industry to automate the recruiting and selecting process, which can remove unconscious human bias that affects the hiring process (Raub, 2018). One of the ideas behind the development of AI in the selection of candidates is to bring higher standards to the selection process independent of the thoughts and beliefs of the interviewer (Miasato and Silva, 2019). Artificial intelligence tools can start with accurate job descriptions and targeted advertisements that match a candidate's skills and abilities to job performance and create a profile of every candidate that indicates which candidate is best suited for the job (Johnson

et al. 2020). In addition, automated resume screening systems allow recruiters to consider more candidates that would be overlooked (Beneduce, 2020). With advances in AI technology, candidate selection becomes impersonal based on data shared with the company and available on the Internet.

*Recruitment efficiency.* HR departments may receive many candidates for every position. Traditional screening and selection that depends on human intervention to evaluate candidate information is the most expensive and discouraging hiring process (Hmoud and Laszlo, 2019). Artificial intelligence can accelerate the hiring procedure, produce an outstanding candidate experience, and reduce costs (Johansson and Herranen, 2019). It can bring job information to applicants faster, allowing them to make informed decisions about their interests early in the hiring process. Artificial intelligence can also screen out many uninterested applicants and remove them from the applicant pool, thus reducing the number of applicants recruiters need to select later. It is even possible to source reticent candidates with the help of artificial intelligence and have more time to concentrate on the best match. Artificial intelligence can not only automate the evaluation of hundreds of resumes on a large scale in a short period, but it can also automatically classify candidates based on the job description provided. Moreover, The final results after the hiring decision can be more easily fed back to the candidate (Raveendra et al., 2020).

*Transactional workload.* The application of AI in recruitment can be described as a "new era in human resources" because artificial intelligence replaces the routine tasks performed by human recruiters, thus changing the traditional practices of the recruitment industry (Upadhyay and Khandelwal, 2018). Most professionals believe that AI is beneficial to recruiters in terms of reducing routine and administrative tasks (Johansson and Herranen, 2019). Recruiters will hand over time-consuming administrative tasks like recruiting, screening, and interviewing to AI, allowing more scope for recruiters to concentrate on strategic affairs (Upadhyay and Khandelwal, 2018).

### Theme II. Why is there algorithmic recruitment discrimination

Algorithms are not inherently discriminatory, and engineers rarely intentionally introduce bias into algorithms. However, bias can still arise in algorithmic recruitment. This issue is closely linked to the fundamental technology behind AI and ML. The ML process can be simplified into several stages, each involving three key components contributing to algorithmic bias: dataset construction, the engineer's target formulation, and feature selection (36KE, 2020). When the dataset lacks diverse representation from different companies, bias may be introduced during the development of algorithmic rules by engineers and when annotators handle unstructured data (Zixun, 2020).

**Datasets: bias soil.** Datasets serve as the foundation of machine learning (ML). If an algorithm's data collection lacks quantity and quality, it will fail to represent reality objectively, leading to inevitable bias in algorithmic decisions. Researchers commonly use a 95% confidence level, which provides 95% certainty but still leaves a one in twenty chance of bias (Raub, 2018). Nearly every ML algorithm relies on biased databases.

One issue arises when datasets are skewed towards accessible and more "mainstream" groups due to the ease of data collection. Consequently, there is an imbalance in the distribution concerning gender and race dimensions (36KE, 2020). If the collected data inadequately represent a particular race or gender, the

resulting system will inevitably overlook or mistreat them in its performance. In the hiring process, insufficient data may exclude historically underrepresented groups (Jackson, 2021). Assessing the success of potential employees based on existing employees perpetuates a bias toward candidates who resemble those already employed (Raghavan et al., 2020).

Without careful planning, most datasets consist of unstructured data acquired through observational measures, lacking rigorous methods in controlled environments (McFarland and McFarland, 2015). This can lead to significant issues with misreporting. When algorithms play a role in decision-making, underrepresented individuals are unequally positioned. Furthermore, as AI improves the algorithm, the model accommodates the lack of representation, reducing sensitivity to the underrepresented groups. The algorithm favors the represented group, operating less effectively for other groups (Njoto, 2020).

Existing social biases are introduced into the dataset. The raw data already reflects social prejudices, and the algorithm also incorporates biased relationships, leading to the “bias in and bias out” phenomenon (36KE, 2020). This phenomenon means that discrimination and disparities exist, just like in forecasting, where historical inequalities are projected into the future and may even be amplified (Mayson, 2018).

A research team at Princeton University discovered that algorithms lack access to the absolute truth. The machine corpus contains biases that closely resemble the implicit biases observed in the human brain. Artificial intelligence has the potential to perpetuate existing patterns of bias and discrimination because these systems are typically trained to replicate the outcomes achieved by human decision-makers (Raso et al. 2018). What is worse, the perception of objectivity surrounding high-tech systems obscures this fact.

In summary, if an algorithmic system is trained on biased and unrepresentative data, it runs the risk of replicating that bias.

**Data feature selection: designer bias.** The introduction of bias is sometimes not immediately apparent in model construction because computer scientists are often not trained to consider social issues in context. It is crucial to make them aware of attribute selection’s impact on the algorithm (Yarger et al., 2019).

The algorithm engineer plays a crucial role in the entire system, from setting goals for machine learning to selecting the appropriate model and determining data characteristics such as labels. If inappropriate goals are set, bias may be introduced from the outset (36KE, 2020).

An engineer is responsible for developing the algorithmic model. If they hold certain beliefs and preconceptions, those personal biases can be transmitted to the machine (Njoto, 2020). Although the device is responsible for selecting employee resumes, it operates based on underlying programming. The programmer guides the AI in making decisions about the best candidate, which can still result in discrimination (Miasato and Silva, 2019).

Furthermore, personal biases can manifest in the selection of data characteristics. For example, engineers may prioritize specific features or variables based on how they want the machine to behave (Miasato and Silva, 2019)). The Amazon hiring case illustrates this, where engineers considered education, occupation, and gender when assigning labels to the algorithm. When gender is considered the crucial criterion, it influences how the algorithm responds to the data.

### **Theme III. Which algorithmic recruitment discrimination exists**

In the recruitment process, algorithmic bias can be manifested in terms of gender, race, color, and personality.

**Gender.** Gender stereotypes have infiltrated the “lexical embedding framework” utilized in natural language processing (NLP) techniques and machine learning (ML). Munson’s research indicates that “occupational picture search outcomes slightly exaggerate gender stereotypes, portraying minority-gender occupations as less professional” (Avery et al., 2023; Kay et al., 2015).

The impact of gender stereotypes on AI hiring poses genuine risks (Beneduce, 2020). In 2014, Amazon developed an ML-based hiring tool, but it exhibited gender bias. The system did not classify candidates neutrally for gender (Miasato and Silva, 2019). The bias stemmed from training the AI system on predominantly male employees’ CVs (Beneduce, 2020). Accordingly, the recruitment algorithm perceived this biased model as indicative of success, resulting in discrimination against female applicants (Langenkamp et al. 2019). The algorithm even downgraded applicants with keywords such as “female” (Faragher, 2019). These findings compelled Amazon to withdraw the tool and develop a new unbiased algorithm. However, this discrimination was inadvertent, revealing the flaws inherent in algorithmic bias that perpetuates existing gender inequalities and social biases (O’neil, 2016).

**Race.** Microsoft’s chatbot Tay learned to produce sexist and racist remarks on Twitter. By interacting with users on the platform, Tay absorbed the natural form of human language, using human tweets as its training data. Unfortunately, the innocent chatbot quickly adopted hate speech targeting women and black individuals. As a result, Microsoft shut down Tay within hours of its release. Research has indicated that when machines passively absorb human biases, they can reflect subconscious biases (Fernández and Fernández, 2019; Ong, 2019). For instance, searches for names associated with Black individuals were more likely to be accompanied by advertisements featuring arrest records, even when no actual records existed. Conversely, searches for names associated with white individuals did not prompt such advertisements (Correll et al., 2007). A study on racial discrimination revealed that candidates with white names received 50% more interview offers than those with African-American names.

**Skin color.** In 2015, Google’s photo application algorithm erroneously labeled a photo of two black people as gorillas (Jackson, 2021). The algorithm was insufficiently trained to recognize images with dark skin tones (Yarger et al., 2023). The company publicly apologized and committed to immediately preventing such errors. However, three years later, Google discontinued its facial identification service, citing the need to address significant technical and policy issues before resuming this service. Similarly, in 2017, an algorithm used for a contactless soap dispenser failed to correctly identify shades of skin color, resulting in the dispenser only responding to white hands and not detecting black and brown ones. These cases serve as examples of algorithmic bias (Jackson, 2021).

**Personality.** The algorithm assesses word choice, tone shifts, and facial expressions (using facial recognition) to determine the candidate’s “personality” and alignment with the company culture (Raso et al., 2018). Notable examples include correlating longer tenure in a specific job with “high creativity” and linking a stronger inclination towards curiosity to a higher likelihood of seeking other opportunities (O’neil, 2016). Additionally, sentiment analysis models are employed to gauge the level of positive or negative emotions conveyed in sentences.

#### **Theme IV. How decreasing algorithmic recruitment discrimination**

Changes should be made at the technical and regulatory levels to ensure that AI algorithms do not replicate existing biases or introduce new ones based on the provided data (Raub, 2018).

##### **Building fair algorithms from a technical perspective**

*Constructing a more unbiased dataset.* Unfair datasets are the root cause of bias. Therefore, a direct approach to addressing algorithmic bias is reconfiguring unbalanced datasets. Using multiple data points can yield more accurate results while carefully eliminating data points that reflect past biases. However, this approach incurs significant costs (Bornstein, 2018).

Another method is to correct data imbalances by using more equitable data sources to ensure fair decision-making (36KE, 2020). Understanding the underlying structure of training data and adjusting the significance of specific data points during training based on known latent distributions makes it possible to uncover hidden biases and remove them automatically. For example, Microsoft revised their dataset for training the Face API, resulting in a 20-fold reduction in the recognition error ratio between men and women with darker skin tones and a 9-fold reduction for women by balancing factors such as skin color, age, and gender (Grabovskyi and Martynovych, 2019).

Integrating “small data” and “big data” can enhance accuracy (36KE, 2020). Data should not solely rely on extensive collections but also focus on precision. While big data analysis tends to emphasize correlations, which can lead to errors when inferring causation, small data, which is more user-specific, offers detailed information and helps avoid such mistakes. Combining the vastness of big data with the precision of small data can help somewhat mitigate hiring errors (Kitchin and Laurialt, 2015).

Biases in datasets can be identified through autonomous testing. The inaccuracies stemming from incomplete past data can be addressed through “oversampling” (Bornstein, 2018). Researchers from MIT demonstrated how an AI system called DB-VEA (unsupervised learning) can automatically reduce bias by re-sampling data. This approach allows the model to learn facial features such as skin color and gender while significantly reducing categorization biases related to race and gender (Amini et al., 2019).

Therefore, constructing a more unbiased dataset is one of the methods that can be employed to tackle algorithmic bias.

*Enhancing algorithmic transparency.* Engineers write algorithmic models, but they often need help understanding the processes that AI undergoes to produce a specific outcome. Many algorithmic biases are difficult to fully understand because their techniques and methods are not easily visible. This leaves many people unaware of why or how they are discriminated against and lacks public accountability (Jackson, 2021). There is an issue of “algorithmic black box” in ML. Therefore, transparency would facilitate remediation when deviant algorithms are discovered and solve the current “black box” dilemma (Shin and Park, 2019).

*Technological tools against bias.* Data blending process. Blendoor is inclusive recruiting and staffing analytics software that mitigates unconscious bias. It takes candidate profiles from existing online job boards and applicant tracking systems to reduce unconscious bias. Blendoor “blends” candidate profiles by removing names, photos, and dates (Yarger et al., 2019). Thus, Blendoor promotes design fairness by assisting underrepresented job seekers and encoding equal opportunity in the algorithm.

Decoupling technique. In resume screening, this technique allows the algorithm to identify the best candidates by

considering variables optimized for other applicants based on specific categories like gender or race rather than the entire applicant pool (Raso et al., 2018). It means that the characteristics selected for minority or female applicants will be determined according to the trends of other minority or female applicants, which may differ from the features identified as successful representatives.

Word embedding. Microsoft researchers found that words exhibit distinct associations in news and web data. For instance, words like “fashion” and “knitting” are more closely related to females, while “hero” and “genius” are more closely related to males (36KE, 2020). Microsoft suggests a simple solution by removing the gender-specific measures in word embedding to reduce “presentation bias.”

Differential testing. Scientists at Columbia University developed Deep Xplore, a software that highlights vulnerabilities in algorithmic neural networks via “coaxing” the system to make mistakes (Xie et al., 2018). Deep Xplore utilizes discrepancy testing, which involves comparing several systems and observing their outputs’ differences. A model is considered vulnerable if all other models consistently predict a particular input while only one model predicts it differently (Gulzar et al., 2019).

Bias detection tool. In September 2018, Google introduced the innovative What-If tool for detecting bias (Mishra, 2022). It assists designers in identifying the causes of misclassification, determining decision boundaries, and detecting algorithmic fairness through interactive visual interfaces. Additionally, Facebook has developed Fairness Flow, an emerging tool for correcting algorithmic bias. Fairness Flow automatically notifies developers if an algorithm makes unfair judgments based on race, gender, or age (Kessing, 2021).

##### **Improving the algorithm’s ethics from a management perspective**

*Internal ethics governance.* Several major technology companies have published AI principles addressing bias governance, signaling the start of self-regulation (36KE, 2020). Microsoft has formed an AI and ethical standards committee to enforce these principles, subjecting all future AI products to ethics scrutiny (Smith and Shum, 2018). Google has responded by introducing a Model Card function, similar to an algorithm manual, that explains the employed algorithm, highlights strengths and weaknesses, and even shares operational results from various datasets (Mitchell et al., 2019).

Algorithmic systems undergo audits to prevent unintended discrimination and make necessary adjustments to ensure fairness (Kim et al., 2021). Regular internal audits allow companies to monitor, identify, and correct biased algorithms. Increased involvement from multiple parties in the data collection process and continuous algorithm monitoring are essential to reduce or eliminate bias (Jackson, 2021). Some companies have introduced AI-HR audits, similar to traditional HR audits, to review employee selection and assess the reliability of AI algorithms and ML data (Yang et al., 2021). Companies should also stay updated on recruitment laws and regulations and ensure compliance with legal requirements.

Considering the programmers behind these algorithms, diversity in the high-tech industry is crucial. Algorithms often reflect the opinions of those who create them. The persistent underrepresentation of women, African-Americans, and Latino professionals in the IT workforce leads to biased algorithms. For instance, a study in 2019 found that only 2.5% of Google’s employees were black, while Microsoft and Facebook had only 4% representation. Another study revealed that 80% of AI professors in 2018 were male. Involving diverse individuals in data collection

and training can regulate and eliminate human bias rooted in algorithms (Jackson, 2021).

Although self-regulation can help reduce discrimination and influence lawmakers, it has potential drawbacks. Self-regulation lacks binding power, necessitating external oversight through third-party testing and the development of AI principles, laws, and regulations by external agencies.

*External supervision.* To ensure transparency and accountability in recruitment, third-party certification and testing of AI products can help mitigate the negative impacts of unreliability. At the “Ethics and Artificial Intelligence” technical conference held at Carnegie Mellon University, the director of Microsoft Research Institute proposed a solution to ensure consistent standards and transparency in AI. Microsoft’s proposal, “Allowing third-party testing and comparison,” aims to uphold the integrity of AI technology in the market (Ong, 2019).

Various organizations have issued principles promoting equity, ethics, and responsibility in AI (Zuiderveen Borgesius, 2020). The Organization for Economic Cooperation and Development (OECD) has provided recommendations on AI, while the European Commission has drafted proposals regarding the influence of algorithmic systems on human rights. In 2019, the European Commission established a high-level expert group on AI, which proposed ethical guidelines and self-regulatory measures regarding AI and ethics.

Public organizations have played a role in establishing mechanisms to safeguard algorithmic fairness. The Algorithm Justice League (AJL) has outlined vital behaviors companies should follow in a signable agreement. Holding accountable those who design and deploy algorithms improves existing algorithms in practice (36KE, 2020). After evaluating IBM’s algorithm, AJL provided feedback, and IBM responded promptly, stating that they would address the identified issue. As a result, IBM significantly improved the accuracy of its algorithm in minority facial identification.

Data protection and non-discrimination laws safeguard against discriminatory practices in algorithmic decision-making. In the EU region, Article 14 of the European Convention on Human Rights (ECHR) guarantees the rights and freedoms outlined in the Convention, prohibiting direct and indirect discrimination (Zuiderveen Borgesius, 2020). Non-discrimination laws, particularly those about indirect discrimination, serve as a means to prevent various forms of algorithmic discrimination. The EU General Data Protection Regulation (GDPR), implemented in May 2018, addresses the impact of ML algorithms and offers a “right to explanation” (e.g., Articles 13–15) (Peña et al., 2020), enabling individuals to request explanations for algorithmic decisions and demand measures to avoid discriminatory influences when handling sensitive data. The GDPR mandates organizations to conduct a Data Protection Impact Assessment (DPIA), with each EU member state must maintain an independent data protection authority vested with investigative powers. Under the GDPR, a data protection authority can access an organization’s premises and computers using personal data (Zuiderveen Borgesius, 2020).

**Investigation and analysis**

Based on Grounded Theory, this section uses a qualitative research approach to explore AI-supported recruitment applications and discrimination.

**Sources and methods**

*Research methodology.* The study is based on Grounded Theory and qualitative analysis of interview data. Glaser and Strauss

**Table 2 Basic information of formal interviewees.**

Number	Age	Gender	Type	Years of work	Role in the workplace
M1	32	Male	Finance	6	Human Resources
M2	25	Male	R&D	2	Technology
M3	31	Male	R&D	6	Technology
M4	36	Male	Education	9	Training
M5	28	Male	Manufacturing	3	Technology
F1	28	Female	Consulting	5	Human Resources
F2	41	Female	Research	16	Technology
F3	36	Female	Education	12	Lectures
F4	55	Female	Finance	32	Counter
F5	27	Female	Retail	2	Sales

(1965,1968) proposed this theory. The basic idea is constructing a theory based on empirical data (Charmaz and Thornberg, 2021). Researchers generally do not make any theoretical assumptions before starting scientific research but start directly from a realistic point of view and summarize several empirical concepts in primary data, which are then raised to systematic theoretical knowledge. Grounded Theory must be supported by empirical evidence, but its most significant characteristic is not its practical nature but the extraction of new ideas from existing facts.

Grounded Theory is a qualitative approach to research that focuses on the importance of “primary sources” (Timmermans and Tavory, 2012). In the study of AI-driven hiring discrimination, the systematic collection and analysis of data are used to uncover intrinsic patterns, construct relevant concepts, and refine relevant theoretical models instead of adopting theoretical assumptions. The current research on the influence factors and measures of AI-driven recruitment discrimination is not intensive enough and lacks corresponding theoretical support. At the same time, Grounded Theory extracts from “primary data” and constructs a theoretical model to study AI-driven recruitment applications and discrimination.

*Interviewees and content*

Participants in the interview: The interview period was June 2023, and the interview form was a face-to-face live/video/telephone interview. The interviewees were selected considering representativeness, authority, and operability, and ten people with experience in recruiting or interviewing with the help of intelligent tools were finally selected for the study. The basic information of the interviewees is shown in Table 2. The study was conducted with the interviewees’ consent. Each interview lasted about 30 min, and notes were taken during the interview. The number of interviewees was determined based on the principle of information saturation.

Before conducting interviews, a large amount of data is collected to understand AI-driven hiring discrimination and propose appropriate improvement strategies. A study of AI-driven hiring discrimination was conducted using “dynamic sampling” and “information saturation” methods.

Interview outline: The interview outline was set in advance around the core objectives of this study, including the following six questions: “Do you know about AI-driven recruitment,” “How do you think about AI-driven recruitment discrimination,” “What do you think is the cause of AI-driven recruitment discrimination,” “Types of AI-driven hiring discrimination,” “Strategies to solve AI-driven hiring discrimination,” and “What other suggestions do you have.” Based on the predefined interview outline, appropriate adjustments were made as the interview progressed.



**Interview ethics:** The interview process is based on three main principles—the right-to-know principle. The interviewer fully understands the purpose, content, and use of the interview before being interviewed, the principle of objectivity. The researcher will guide the respondent to ask questions and answer what they cannot understand. Respondents make objective statements of their willingness that are not influenced by external factors, the principle of confidentiality. Interviews will be conducted anonymously, and the personal information of the interviewees will not be disclosed. The interviewee's privacy is fully respected, and the original data is replaced by figures, which will be used only for the interviewer's reference and analysis, appropriately kept by the interviewer, and used only for this study and no other purpose.

**Interview tools:** Nvivo 12.0 Plus qualitative analysis software was used as an auxiliary tool to clarify ideas and improve work efficiency.

**Data organization after the interview:** Within two working days after the completion of the interview, the analysis and organization of the interview data was completed. The Nvivo 12 plus software coded the interview data in three layers from the bottom up, with the content as the center.

The first layer was open coding. The interview data of 10 interviewees were imported and, using the software, parsed word by word to clarify the meaning of words and sentences, give an interpretation of the data, and obtain free nodes. The data from each section was summarized and inferred to organize the interviewees' perceptions of AI-driven recruitment, and each node was given a name to derive the first-level nodes.

Next, the second-level spindles were coded. The researchers unfolded the induction of interrelated classes for the nodes formed by the open coding, constructed the relationship between concepts and classes, and coded a spindle concept, which, after the spindle coding, would form a second-level node.

The third part is the three-level core coding. Another coding core class genus is selected based on the secondary spindle, and a tertiary code is developed.

**Interview quality control:** In order to ensure the credibility of the interview results, the method adopted in this study is to use a uniform way of asking questions to different interviewees. Each round of interviews should be kept between 20 and 30 min. Too long an interval will reduce the effectiveness of feedback on the questions. Also, interviews should not span more than one month to ensure the timeliness of the information obtained.

**Results.** Applying the Nvivo12 Plus qualitative analysis method, 182 free nodes were obtained by three-level coding, 31 primary nodes were formed after analogy, and 11 secondary nodes were deduced inductively. Finally, five core genera of tertiary nodes were identified (see Table 3).

**Open level 1 coding.** The interviews with 10 respondents resulted in 182 words and sentences related to AI-driven recruitment applications and discrimination, which were conceptualized and merged to form 31 open-ended Level 1 codes.

**Main-axis second-level coding.** The spindle codes were analyzed through cluster analysis to analyze the correlation and logical order among the primary open codes, forming more generalized categories. Eleven spindle codes were extracted and summarized.

**Core-type tertiary coding.** The Grounded Theory steps resulted in 31 open primary and 11 secondary spindle codes. Further categorization and analysis revealed that when "AI-driven

recruitment applications and discrimination" is used as the core category, the five main categories are AI-driven recruitment applications, AI-driven recruitment effects, causes of AI-driven recruitment discrimination, types of AI-driven recruitment discrimination, and AI-driven recruitment discrimination measures.

The coding process described above was exemplified by an interview with a researcher, F2, who had taught information science at a university for 2 years and was now employed at an intelligent technology R&D company. After the interview, F1's information was analyzed and explored in a three-level coding process.

Under the three-level node AI-driven recruitment application, F2 suggested that the AI tools currently developed could assist companies with simple recruitment tasks, including online profile retrieval, analysis, and evaluation. However, this technical engineer suggested that candidate assessment for high-level positions suits human-machine collaboration, although machines have an advantage in candidate profile searches.

Supported by three-level nodes of AI-driven recruiting effectiveness, F2 suggests that machine applications in recruiting can relieve human transactional workload, and chatbot Q&A services improve recruiting efficiency.

In the context of the causes of AI-driven hiring discrimination at the third level, F2 suggests that some job seekers are unfamiliar with the hiring interface and how to use it, leading to unfair interviews. She suggested the need for organizers to prepare usage guidelines or mock interview exercises. She argues that much of the data needed for intelligent machine learning comes from internal companies or external market supplies and that this data lacks fair scrutiny. It is even possible that the source data fed into the machines is compromised.

Under the tertiary node AI-driven hiring discrimination, F2 is concerned that the machines may misevaluate candidates due to individual differences, such as intelligence, or external characteristics, such as skin color. Moreover, some discriminatory judgments are difficult to resolve under current technology.

Under the tertiary node AI-driven hiring discrimination measures, F2 proposes utilizing technical tools, such as learning impartial historical data, or non-technical tools, such as anti-AI discrimination laws. She argues that in the future, humans use AI tools to solve more complex decisions, not just limited to hiring. Instead, humans need to embrace and accept the widespread use of machines.

Synthesizing the above analysis, the final overview of the AI-driven recruitment application and discrimination framework is obtained (see Fig. 3). After the conceptual model was constructed, the remaining original information was coded and comparatively analyzed, and no new codes were generated, indicating that this study was saturated.

**Discussion.** An analysis of interview results conducted using Grounded Theory indicates that AI-supported hiring discrimination should be approached from five perspectives. These perspectives align with the thematic directions identified through our literature review.

Firstly, AI-driven hiring applications impact various aspects, such as reviewing applicant profiles online, analyzing applicant information, scoring assessments based on hiring criteria, and generating preliminary rankings automatically.

Secondly, interviewers perceive benefits in AI-driven recruitment for job seekers. It eliminates subjective human bias, facilitates automated matchmaking between individuals and positions, and provides automated response services. Moreover, AI reduces the workload on humans and enhances efficiency.

**Table 3 Summary of nodes and reference points at all levels after coding of interview data.**

Level 3 nodes	Level 2 nodes	Level 1 nodes	Number of sources	Number of coding reference points
AI-driven recruitment application	Sourcing	Online review of job applicants	4	4
		Analysis of personal data	3	4
		Assessing against eligibility metrics	3	3
		Automated scoring and ranking of applicants	6	10
	Schedule interviews	Automatically schedule interview process	4	5
		Analyze the candidate's expressions, reactions	4	7
		Automatically respond to questions from candidates	8	11
	Select	Evaluate the available compensation package	3	5
		Predict candidate performance	5	6
	Effectiveness of AI-driven recruitment	Benefit job seekers	Eliminate human bias	7
Q&A with chatbots			9	9
Automatically matching the right job			5	6
Beneficial to organizations		Reducing labor costs	5	7
		Assists with transactional tasks	3	3
		Improves efficiency	5	6
Causes of AI-driven hiring discrimination	Caused by AI software	Flawed software program design	1	1
		Historical data used unfairly	4	9
		Inaccessible software application interface	3	3
	User-induced	Lack of familiarity or access to manuals	2	2
		Lack of training in machine application affects the interview	7	8
		Simulated machine Q&A to obtain high scores	4	6
Types of AI-driven hiring discrimination	Extrinsic factors	Gender	8	9
		Nationality	3	3
	Intrinsic factors	Personality	4	6
		IQ	2	3
Anti-discrimination measures	Technical tools	Fair machine programs	8	9
		Software application guidance	7	7
		Machine learning non-discrimination data	3	4
	Non-technical means	Government oversight	5	7
		Laws related to AI-driven hiring	6	9
		Third-party audits	1	2

Thirdly, concerns are raised regarding potential hiring discrimination perpetrated by machines. This can arise from AI tools, such as partial source data, or users unfamiliar with user interfaces and operations.

Fourthly, intrinsic factors like personality and IQ, as well as extrinsic factors like gender and nationality, have been observed to influence the accurate identification and judgment of AI systems concerning hiring discrimination.

Fifthly, respondents offer recommendations for combating discrimination by machines, including technical and non-technical approaches.

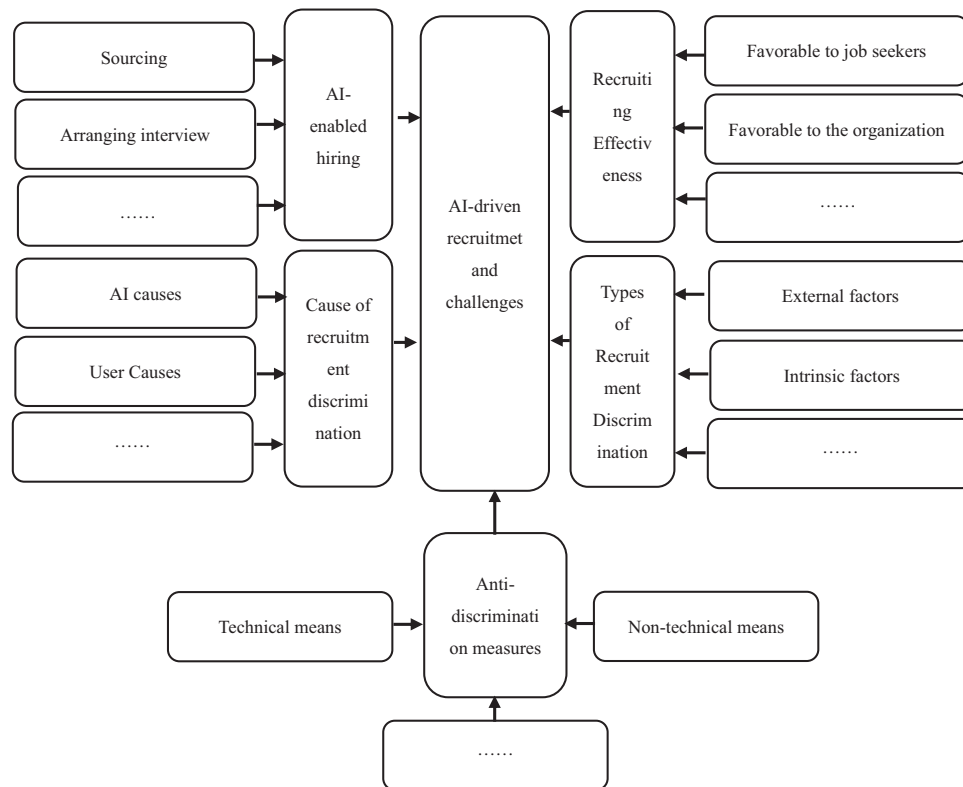
**Recommendations for future studies**

This study conducted a literature review to analyze algorithmic recruitment discrimination's causes, types, and solutions. Future research on algorithmic recruitment discrimination could explore quantitative analysis or experimental methods across different countries and cultures. Additionally, future studies could examine the mechanics of algorithmic recruitment and the technical rules that impact the hiring process. It would be interesting to analyze the psychological effects of applying this algorithmic recruitment

technique on various populations (gender, age, education level) from an organizational behavior perspective. While recent studies have primarily discussed discrimination theory in the traditional economy's hiring market, future theoretical research should consider how advanced technology affects equity in hiring within the digital economy.

**Conclusion**

The study concludes that the fourth industrial revolution introduced technological innovations significantly affecting the recruitment industry. It extends the analysis of statistical discrimination theory in the digital age and adopts a literature review approach to explore four themes related to AI-based recruitment. The study argues that algorithmic bias remains an issue while AI recruitment tools offer benefits such as improved recruitment quality, cost reduction, and increased efficiency. Recruitment algorithms' bias is evident in gender, race, color, and personality. The primary source of algorithmic bias lies in partial historical data. The personal preferences of algorithm engineers also contribute to algorithmic bias. Technical measures like constructing unbiased datasets and enhancing



**Fig. 3 AI-driven recruitment applications and discrimination.** The comprehensive analysis produces an overview of AI-driven recruitment applications and recognition frameworks.

algorithm transparency can be implemented to tackle algorithmic hiring discrimination. However, strengthening management measures, such as corporate ethics and external oversight, is equally important.

**Data availability**

The study is still ongoing, and the results of subsequent analyses will continue to be applied to valuable and critical projects. Relevant data are currently available only to scholars conducting similar research, with the prerequisite of signing a confidentiality agreement. Corresponding author can be contacted for any requests.

Received: 18 February 2023; Accepted: 29 August 2023;  
Published online: 13 September 2023

**References**

36KE (2020) From sexism to recruitment injustice, how to make AI fair to treat. <https://baijiahao.baidu.com/s?id=1663381718970013977&wfr=spider&for=pc>  
 Ahmed O (2018) Artificial intelligence in HR. *Int J Res Anal Rev* 5(4):971–978  
 Albert ET (2019) AI in talent acquisition: a review of AI-applications used in recruitment and selection. *Strateg HR Rev* 18(5):215–221. <https://doi.org/10.1108/SHR-04-2019-0024>  
 Allal-Chérif O, Yela Aránega A, Castaño Sánchez R (2021) Intelligent recruitment: how to identify, select, and retain talents from around the world using artificial intelligence. *Technol Forecast Soc Change* 169:120822. <https://doi.org/10.1016/j.techfore.2021.120822>  
 Amini A, Soleimany AP, Schwarting W, Bhatia SN, Rus D (2019) Uncovering and mitigating algorithmic bias through learned latent structure. In: Conitzer V, Hadfield G, Vallor S (eds) Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society. Association for Computing Machinery

Avery M, Leibbrandt A, Vecchi J (2023) Does artificial intelligence help or hurt gender diversity? In: Conitzer V, Hadfield G, Vallor SE (eds) *vidence from two field experiments on recruitment in Tech*, 14 February 2023. Association for Computing Machinery  
 Beattie G, Johnson PJPP, Education PiH (2012) Possible unconscious bias in recruitment and promotion and the need to promote equality. *Perspect: Policy Pract High Educ* 16(1), 7–13  
 Beneduce G (2020) Artificial intelligence in recruitment: just because it’s biased, does it mean it’s bad? NOVA—School of Business and Economics  
 Black JS, van Esch P (2020) AI-enabled recruiting: what is it and how should a manager use it? *Bus Horiz* 63(2):215–226. <https://doi.org/10.1016/j.bushor.2019.12.001>  
 Bogen M, Rieke A (2018) Help wanted: an examination of hiring algorithms, equity, and bias. *Upturn*  
 Bornstein S (2018) Antidiscriminatory algorithms. *Alabama Law Rev* 70:519  
 Cain GG (1986) The economic analysis of labor market discrimination: a survey. *Handb Labor Econ* 1:693–785  
 Charmaz K, Thornberg R (2021) The pursuit of quality in grounded theory. *Qual Res Psychol* 18(3):305–327. <https://doi.org/10.1080/14780887.2020.1780357>  
 Chen Z (2022) Artificial intelligence-virtual trainer: innovative didactics aimed at personalized training needs. *J Knowl Econ*. <https://doi.org/10.1007/s13132-022-00985-0>  
 Chen Z (2023) Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. *Cogn Technol Work* 25(1):135–149  
 Correll SJ, Benard S, Paik I (2007) Getting a job: is there a motherhood penalty? *Am J Sociol* 112(5):1297–1338  
 Dickinson DL, Oaxaca RL (2009) Statistical discrimination in labor markets: an experimental analysis. *South Econ J* 76(1):16–31  
 Faragher JJEWSTBANHS (2019) Is AI the enemy of diversity? *People Management*  
 Fernández C, Fernández A (2019) Ethical and legal implications of AI recruiting software. *Ercim News* 116:22–23  
 Grabovskiy V, Martynovych O (2019) Facial recognition with using of the microsoft face API Service. *Electron Inf Technol* 12:36–48  
 Gulzar MA, Zhu Y, Han X (2019) Perception and practices of differential testing. In: Conitzer V, Hadfield G, Vallor S (eds) 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). Association for Computing Machinery  
 Hmoud B, Laszlo V (2019) Will artificial intelligence take over human resources recruitment and selection? *Netw Intell Stud* 7(13):21–30

- Jackson MC (2021) Artificial intelligence & algorithmic bias: the issues with technology reflecting history & humans. *J Bus Technol Law* 16:299
- Johansson J, Herranen S (2019) The application of artificial intelligence (AI) in human resource management: current state of AI and its impact on the traditional recruitment process. Bachelor thesis, Jonkoping University
- Johnson RD, Stone DL, Lukaszewski KM (2020) The benefits of eHRM and AI for talent acquisition. *J Tour Futur* 7(1):40–52
- Kay M, Matuszek C, Munson SA (2015) Unequal representation and gender stereotypes in image search results for occupations. In: Conitzer V, Hadfield G, Vallor S (eds) Proceedings of the 33rd annual ACM conference on human factors in computing systems. Association for Computing Machinery
- Kessing M (2021) Fairness in AI: discussion of a unified approach to ensure responsible AI development. Independent thesis Advanced level, KTH, School of Electrical Engineering and Computer Science (EECS)
- Kim PT, Bodie MTJJOL, Law E (2021) Artificial intelligence and the challenges of workplace. *Discrim Privy* 35(2):289–315
- Kitchin R, Lauriault TPJG (2015) Small data in the era of big data 80(4):463–475
- Köchling A, Wehner MC, Warkocz J (2022) Can I show my skills? Affective responses to artificial intelligence in the recruitment process. *Rev Manag Sci*. <https://doi.org/10.1007/s11846-021-00514-4>
- Langenkamp M, Costa A, Cheung C (2019) Hiring fairly in the age of algorithms. Available at: SSRN 3723046
- Lloyd K (2018) Bias amplification in artificial intelligence systems. arXiv preprint arXiv:07842
- Lundberg SJ, Startz R (1983) Private discrimination and social intervention in competitive labor market. *Am Econ Rev* 73(3):340–347
- Mayson SG (2018) Bias in, bias out. *Yale Law J* 128(8):2218–2300
- McFarland DA, McFarland HR (2015) Big data and the danger of being precisely inaccurate. *Big Data Soc Hum Resour Manag* 2(2):2053951715602495
- Miasato A, Silva FR (2019) Artificial intelligence as an instrument of discrimination in workforce recruitment. *Acta Univ Sapientiae: Legal Stud* 8(2):191–212
- Mishra P (2022) AI model fairness using a what-if scenario. In: Practical explainable AI using Python. Springer, pp. 229–242
- Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, ... Gebru T (2019) Model cards for model reporting. In: Conitzer V, Hadfield G, Vallor S (eds) Proceedings of the conference on fairness, accountability, and transparency. Association for Computing Machinery
- Newell S (2015) Recruitment and selection. Managing human resources: personnel management in transition. Blackwell Publishing, Oxford
- Njoto S (2020) Research paper gendered bots? Bias in the use of artificial intelligence in recruitment
- O'neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Crown
- Ong JH (2019) Ethics of artificial intelligence recruitment systems at the United States Secret Service. <https://scholarship.library.jhu.edu/handle/1774.2/61791>
- Peña A, Serna I, Morales A, Fierrez J (2020) Bias in multimodal AI: testbed for fair automatic recruitment. In: Conitzer V, Hadfield G, Vallor S (eds) Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. Association for Computing Machinery
- Prince AE, Schwarcz D (2019) Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Rev* 105:1257
- Raghavan M, Barocas S, Kleinberg J, Levy K (2020) Mitigating bias in algorithmic hiring: evaluating claims and practices. In: Conitzer V, Hadfield G, Vallor S (eds) Proceedings of the 2020 conference on fairness, accountability, and transparency. Association for Computing Machinery
- Raso FA, Hillgoss H, Krishnamurthy V, Bavitz C, Kim L (2018). Artificial intelligence & human rights: opportunities & risks. Available at: SSRN3259344 (2018-6)
- Raub M (2018) Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Ark Law Rev* 71:529
- Raveendra P, Satish Y, Singh P (2020) Changing landscape of recruitment industry: a study on the impact of artificial intelligence on eliminating hiring bias from recruitment and selection process. *J Comput Theor Nanosci* 17(9):4404–4407
- Ruwanpura KN (2008) Multiple identities, multiple-discrimination: a critical review. *Fem Econ* 14(3):77–105
- Samuelson PA (1952) Spatial price equilibrium and linear programming. *Am Econ Rev* 42(3):283–303
- Shaw J (2019) Artificial intelligence and ethics. *Perspect: Policy Pract High Educ* 30, 1–11
- Shin D, Park YJCIHB (2019) Role of fairness, accountability, and transparency in algorithmic affordance. *Perspect: Policy Pract High Educ* 98, 277–284
- Smith B, Shum H (2018). The future computed. Microsoft
- Tilcsik A (2021) Statistical discrimination and the rationalization of stereotypes. *Am Sociol Rev* 86(1):93–122
- Timmermans S, Tavory I (2012) Theory construction in qualitative research: from grounded theory to abductive analysis. *Sociol Theory* 30(3):167–186. <https://doi.org/10.1177/0735275112457914>
- Upadhyay AK, Khandelwal K (2018) Applying artificial intelligence: implications for recruitment. *Strateg HR Rev* 17(5):255–258
- van Esch P, Black JS, Ferolie J (2019) Marketing AI recruitment: the next phase in job application and selection. *Comput Hum Behav* 90:215–222. <https://doi.org/10.1016/j.chb.2018.09.009>
- Xie X, Ma L, Juefei-Xu F, Chen H, Xue M, Li B, ... See S (2018) Deephunter: Hunting deep neural network defects via coverage-guided fuzzing. Available at: arXiv preprint arXiv:01266
- Yang J, Im M, Choi S, Kim J, Ko DH (2021) Artificial intelligence-based hiring: an exploratory study of hiring market reactions. *Japan Labor Issues* 5(32):41–55
- Yarger L, Payton FC, Neupane B (2019) Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Inf Rev* 44(2):383–395
- Yarger L, Smith C, Nedd A (2023) 11. We cannot build equitable artificial intelligence hiring systems without the inclusion of minoritized technology workers. In: Conitzer V, Hadfield G, Vallor S (eds) Handbook of gender and technology: environment, identity, individual. p. 200. Association for Computing Machinery
- Zhang J, Chen Z (2023) Exploring human resource management digital transformation in the Digital Age. *J Knowl Econ*. <https://doi.org/10.1007/s13132-023-01214-y>
- Zixun L (2020) From sexism to unfair hiring, how can AI treat people fairly? <https://baijiahao.baidu.com/s?id=1662393382040525886&wfr=spider&for=pc>
- Zuiderveen Borgesius FJ (2020) Strengthening legal protection against discrimination by algorithms and artificial intelligence. *Int J Hum Rights* 24(10):1572–1593

### Author contributions

ZSC conceived, wrote, and approved the manuscript.

### Competing interests

The author declares no competing interests.

### Ethical approval

Approval was obtained from the ethics committee of the University NUAA. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

### Informed consent

Each participant in this study willingly provided informed permission after being fully told of the study's purpose, its methods, its participants' rights, and any possible dangers. They were therefore assured of their comprehension and consent.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-02079-x>.

**Correspondence** and requests for materials should be addressed to Zhisheng Chen.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons

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

© The Author(s) 2023