# CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Security Studies


# Algorithmic Discrimination: An Ethical Analysis of Algorithmic Bias in Employment and Hiring Practices

**2024** **Joana Silva**

# CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Security Studies


# Algorithmic Discrimination: An Ethical Analysis of Algorithmic Bias in Employment and Hiring Practices

Master's Thesis


Author of the Thesis: Joana dos Santos Silva

Study programme: International Security Studies

Supervisor: doc. PhDr. Vít Střítecký, M.Phil., Ph.D.



Year of the defence: 2024

## Declaration

The author hereby declares that she compiled this thesis using only the listed literature and resources.
The author hereby declares that the thesis has not been used to obtain any other academic title.
This research was conducted following the ethical guidelines and standards of Charles University.
The author grants Charles University permission to use this thesis in whole or in part for study and scientific purposes.

Prague, 3rd July 2024 Joana Silva

## References

SILVA, Joana. *Algorithmic Discrimination: An Ethical Analysis of Algorithmic Bias in Employment and Hiring Practices.* Praha, 2024. Master's thesis (Mgr). Charles University, Faculty of Social Sciences, Institute of Political Studies, Department of Security Studies. Supervisor doc. PhDr. Vít Střítecký, M.Phil., Ph.D.

# Table of Contents

# Introduction

In a world of constant innovation and profound transformations, Artificial Intelligence (AI) has become a central and omnipresent element of the digital and technological revolution we are witnessing.

It is crucial that we deconstruct the basic terms of what algorithmic discrimination and bias are, as well as the notions of transparency, fairness, and justice that should be reflected in any technological advancement.

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in our society has profoundly impacted numerous aspects of human existence. The prominence and significant contribution provided by AI tools in executing and optimizing various daily tasks, habits, behaviors, and even ways of functioning or thinking, is undeniable (Mayer-Schönberger and Cukier, 2013).

Algorithms have brought about significant changes in our lives, some in more explosive ways, others more subtly. They have come to revolutionize the modern industry, not only to answer objective questions but also to optimize subjective, complex issues that involve value judgments, such as in the employment sector and recruitment practices traditionally performed by human resources professionals (Zuboff, 2019).

The adoption of algorithmic decision-making tools expands a world of potential, including objectivity, efficiency, and cost reduction, streamlining hiring processes, and identifying the ideal candidate. However, like any technological advancement, it carries concerns. These systems can often inherit biases present in their primary design data or even develop new biases through flawed design and implementation, raising significant ethical reflections on the perpetuation of biases and algorithmic discrimination, especially against certain minority demographic groups (Noble, 2018).

This thesis aims to contribute to the constructive dialogue on the prevalence of algorithmic bias in employment and recruitment practices, exploring and conducting an analysis of the ethical dimension of existing biases. Given the increasing reliance on these technologies, there is an urgent need to investigate and understand, through real case studies of algorithmic bias, how these biases manifest in recruitment systems, the underlying causes of algorithmic bias, the ethical considerations inherent in these selection and recruitment practices, and their impact, as well as propose recommendations and potential mitigation strategies that promote justice, equity, and transparency in algorithmic decision-making processes.

As AI technologies become increasingly integrated into decision-making processes, it is essential to ensure that these systems not only function but do so fairly and do not exacerbate existing social inequalities (Benjamim, 2019).

This research is crucial as it addresses a gap in the current literature. While there is already extensive discussion on algorithmic bias and the inherent prejudices in AI, less attention has been given to the more specific challenges and ethical considerations in the workplace context, including recruitment processes. This study intends to bridge this gap by analyzing the ethical issues inherent in this problem and presenting possible recommendations and practical strategies for professionals and policymakers.

It is imperative to critically assess the ethical dimensions of algorithmic discrimination in these processes. To this end, we will analyze these dimensions through the lens of two fundamental theories that provide a holistic and differentiated understanding of the ethical issues at play: Critical Race Theory (CRT) and Intersectionality Theory. On one hand, Critical Race Theory offers insight into the role of systemic inequalities embedded in current societal structures, especially those mediated by technology (Delgado & Stefancic, 2017), and how these inequalities can persist in contexts that are ostensibly neutral and objective. On the other hand, through Intersectionality Theory, we gain a perspective on how various often marginalized social identities, through race and gender, can intersect and culminate in exacerbated situations of discrimination and disadvantage (Crenshaw, 1989). In light of this framework, we aim to provide a comprehensive ethical framework to better understand and analyze algorithmic discrimination in recruitment processes.

Regarding the applied methodology, we adopted a qualitative methodology, analyzing real case studies of algorithmic bias in employment. This choice as a primary data source is strategic, allowing for an in-depth analysis of specific documented cases of algorithmic discrimination. This methodological approach allows us to deeply explore ethical issues, providing a robust analytical framework for qualitative data patterns, and enabling a holistic understanding of how algorithmic bias manifests and spreads in these decision-making processes. We will collect data through a systematic review of the literature, reports, media, and through public policies of companies that will be the focus of this study. In light of these well-documented cases, we will identify, such as Amazon's[1] AI recruitment tool (Dastin, 2018), gig economy[2] platforms like TaskRabbit and Fiverr (Edelman, Luca & Svirsky, 2017), and HireVue[3], a specialist in video interview analysis through AI. These case studies not only provide empirical evidence of the existence and

[1] Amazon is a major player in the global e-commerce market and technology sector, recognized as one of the "Big Five" technology companies, a group that also includes Google, Apple, Microsoft, and Facebook (Satariano, 2020).
[2] The Gig Economy operates based on platforms that serve as an intermediary between the service provider and the end customer, usually are characterized by short-term, flexible and often precarious work. (Kalleberg & Vallas, 2018; De Stefano, 2016).
[3] HireVue is a company specialized in video interviewing technology and AI-driven recruitment solutions (Hirevue, 2021).

inherent impact of algorithmic bias but also help us deeply understand and analyze the mechanisms through which these biases operate.

Through thematic analysis in light of these cases, we aim to understand the patterns of the underlying causes of algorithmic discrimination, namely its primary failures in data and algorithm design, understanding, and implementation, and to follow the process that will lead us to potential mitigation strategies that promote justice, equity, and transparency in these algorithmic processes, allowing technological innovation to progress alongside ethical responsibility, and ensuring that the full potential of AI technologies is harnessed for positive social change.

# 2. Background and Literature Review

The integration of automation into decision-making processes has transformed numerous aspects of societal life, particularly in hiring and recruitment practices, which are the focus of this study. In this literature review section, we will examine existing research on algorithmic discrimination, focusing on three main areas: the nature of bias, inherent causes, and ethical repercussions in society.

First, we will attempt to deconstruct the concept of algorithmic discrimination and the foundational concepts of justice, equity, and transparency in the algorithmic field. We will also discuss and understand the main causes and types of algorithmic bias, thus laying a comprehensive and solid foundation for discussing and understanding the complexities of algorithmic discrimination in recruitment processes, as well as for developing strategies to mitigate its impact. Algorithms themselves are not inherently biased; they simply follow a set of instructions that reflect the data they are trained on, and the choices made by those who train them (Binns, 2018). In this sense, this discussion is fundamental to working towards harnessing the full technological and innovative potential of algorithmic decision-making while striving for justice, transparency, and equity throughout the use of AI algorithms.

## 2.1 The Nature of Algorithmic Discrimination, Causes, and Ethical Implications

First, it is essential to define the basic concepts of algorithmic discrimination, which will be the focus throughout this study, as well as the concepts we aim to achieve during the recruitment and hiring practices: justice, transparency, and equity.

Each of the concepts—justice, transparency, and equity—is crucial and plays pivotal roles both collectively and individually in the technological context, and in this case, in the ethical analysis of decision-making processes based on algorithms. When we briefly discuss the concept of algorithmic justice, it relates to the equitable treatment by algorithms, meaning, in this specific case, that the decisions made during the process should be fair and the system should not disproportionately disadvantage candidates from certain demographic groups (Barocas & Selbst, 2016). When we talk about algorithmic transparency, we aim for the entire process to be as clear as possible for all parties involved. This means that the way data is collected, the design of the algorithm, its operation, implementation, and especially the understanding of the decision-making process should be accessible and clear (Diakopoulos, 2016). Finally, the crucial concept of accountability in algorithm-based decision-making processes. This refers to the responsibility of organizations and users of these models to ensure they are used correctly and ethically, without any bias in their results. It is extremely important that stakeholders have the capacity, beyond understanding the processes, to be held accountable if the systems and mechanisms negatively impact individuals (Pasquale, 2015).

Algorithmic discrimination occurs when automated decision-making systems produce biased outcomes, perpetuate existing prejudices, and may even exacerbate them, as O'Neil (2016) notes, based on characteristics such as gender, age, race, or socioeconomic status. This creates new forms of inequality among individuals, particularly in areas that significantly impact their lives and equality of opportunity. These biased results disproportionately affect certain groups, usually minorities, often without the explicit intention of the users of this technology, who primarily seek efficiency and objectivity in decision-making processes.

This concept encompasses not only all the biases that arise throughout the entire technical process—be it in data collection, feature selection, algorithm design, model training, or result interpretation (Barocas & Selbst, 2016)—but also highlights issues deeply rooted in societal structures and norms (Noble, 2018).

The problem with the quality of the data used by algorithms may originate from the algorithm creators themselves or the database used by the AI. In the first case, AI programmers, when creating an algorithmic model, initially select the information that will be made available to the software that will analyze or make decisions. This means that bias can arise from flaws in programming or in the execution of the task for which the algorithm was designed, including reflecting the programmer's subjectivities, such as their social context, personal beliefs, emotions, prejudices, etc.

All the small parts of this technical process are extremely important, as algorithmic discrimination can arise in any of them, from various sources, often interrelated.

The causes of algorithmic bias are multifaceted. One of the most common causes of algorithmic discrimination is related to data collection, specifically the use of compromised historical training data, which can be biased. Algorithms, particularly those relying on machine learning, depend heavily on historical data patterns that may inherently contain implicit biases. If previous recruitment processes, on which the data collection is based, reflect favoritism towards certain groups over others—such as favoring male candidates for leadership roles over female candidates—these biases can be encoded into the algorithm as patterns and perpetuated, leading to discriminatory outcomes (O'Neil, 2016). Hanna et al. (2020), in their article "Towards a Critical Race Methodology in Algorithmic Fairness," demonstrate that without a thorough analysis of the social and economic context in which the data is produced, algorithms will inevitably reproduce existing inequalities. The data, which should ostensibly be neutral, end up reproducing systemic and historical biases quite significantly.

This issue can be further exacerbated when training data does not represent the diversity of the population, resulting in sample bias, where algorithms disproportionately favor the most represented groups (Eubanks, 2018). This type of bias, sample bias, is especially problematic in algorithmic discrimination within hiring and recruitment practices, as it leads to underrepresentation of certain demographic data and a lack of reflection in final hiring outcomes. For example, if an algorithm is trained on resumes from a particular demographic group, such as adult men from a specific industry or university, it may overlook candidates from other backgrounds and demographic groups, thereby excluding the potential of underrepresented groups (Kim, 2017). In the context of the job market, it's important to keep in mind that in an increasingly globalized world, these issues are increasingly concerning. Companies seek to introduce a more diverse and competent workforce, so the importance of training algorithms is crucial for the hiring process to be as fair and impartial as possible (Raji & Buolamwini). By acknowledging this sample bias, various stakeholders using artificial intelligence in hiring and recruitment systems can work towards creating more inclusive hiring practices that reflect the desired workforce diversity.

Besides the variables indicated in data collection, other factors in different phases of the process, previously mentioned, can significantly lead to algorithmic discrimination, such as the selection of characteristics in the algorithms. This selection of variables must be meticulously chosen because, as Noble (2018) and Barocas & Hardt (2019) point out, seemingly neutral characteristics can serve as proxies for more sensitive variables. For example, standardizing an algorithm to include the candidates' postal codes as one of the variables in the selection process. This neutral variable could serve as a means for indirect discrimination, correlated with protected characteristics based on race or socioeconomic status, as the geographic location is often correlated with these factors, potentially leading to inadvertently discriminatory outcomes (Noble, 2018).

Finally, regarding the implementation and interpretation phase, even when an algorithm is well-designed and trained from its inception with fully impartial and equitable data, the way the results are used by companies, organizations, or policymakers can introduce biases and lead to skewed outcomes. If users of these automated recruitment methods rely too heavily on the algorithm's recommendations without considering any context or limitations, they may inadvertently reject qualified candidates who do not fit the algorithmic profile. This algorithmic bias, known as automation bias, can lead to an over-reliance on the computed results, underestimating the ultimate value of human judgment, as argued by Cummings (2004). Beyond this excessive reliance on algorithmic interpretation, the interpretation of results is also hampered by the opacity of the algorithms, which complicates understanding of their function and decision-making processes, generating distrust and questions when this process is not clear and transparent (Doshi-Velez & Kim, 2017).

Lastly, the norms and culture of recruiters and their respective companies can influence how algorithms are interpreted and implemented. If there is a marked lack of commitment to inclusion and diversity within these organizations, the algorithms, regardless of their design and conception, will consequently reinforce existing biases (Binns, 2018).

It is observed that the starting point of the software always comes from the programmer. However, during the development and creation of an algorithmic model, algorithms can learn automatically from the pre-selected database, bypassing human intervention in controlling the stages until the end of the analysis or decision-making process. As Barocas, Hardt, and Narayanan (2019) argue, biases in data collection, labeling, and even in the design of the algorithmic model itself can lead to discriminatory outcomes. However, it must be emphasized that the fairness of these systems should not be isolated solely as a technical issue but should also be considered within a broader and more holistic socio-technical context.

# 3. Theoretical Framework

To conduct a deep ethical analysis of algorithmic discrimination in employment and recruitment practices, it is crucial to establish a robust and holistic theoretical framework that illuminates the complexities of our topic. This section explores various theoretical perspectives that can provide a differentiated understanding of the complex ethical issues of this context.

Integrating Critical Race Theory (CRT) into the core of this debate highlights that racism is not merely a product of individual prejudices but is deeply embedded in social, political, and legal structures, often perpetuating inequality and marginalization (Crenshaw et al., 1995). In the context of the current technological society and the focus of this study, algorithmic discrimination in employment and recruitment practices, CRT helps elucidate how biases embedded in data and algorithmic processes can not only reproduce these biases in automated decision-making processes but also reinforce pre-existing systemic social stratifications. This theory provides a critical lens to examine the role of systemic inequalities in society, particularly those mediated by technology (Delgado & Stefancic, 2017), and helps understand how historical and systemic prejudices can be encoded into algorithms through data and automated decision-making mechanisms using Artificial Intelligence (AI).

CRT also offers a perspective on the lived experiences of marginalized minority groups and the perspectives of those affected by algorithmic bias. This approach is crucial as it seeks to involve the affected communities, ensuring that their perceptions are considered during algorithmic processes, which is vital for developing more equitable technologies that aim to achieve a deeper understanding of the impact of these systems in the real world (Delgado & Stefancic, 2017).

CRT further provides a critical view of what Derrick Bell (1980) terms "interest convergence," arguing that genuine justice and racial progress only occur when the interests of dominant groups are at stake and that social change for minorities only truly happens when it aligns with the interests of the dominant majority. Applying this perspective to our topic, it becomes clear that companies, organizations, and policymakers must recognize the necessity and clear benefit of making every effort to understand algorithmic systems to mitigate any form of algorithmic bias in recruitment and hiring practices, even when algorithms appear neutral and equitable at first glance.

The other primary theoretical dimension we intend to address is Intersectionality, initially introduced by Kimberlé Crenshaw (1989). This theory provides fundamental insights into understanding the complex interplay between various forms of discrimination, essential for analyzing how algorithmic bias may occur in the workplace and in recruitment practices through algorithmic systems. Intersectionality is crucial for examining algorithmic discrimination in decision-making processes in employment and automated recruitment practices, as it offers a deep understanding that individuals can face discrimination in multiple ways, with different forms intersecting and exacerbating discrimination. Crenshaw (1989) argues that individuals do not suffer discrimination based on a single identity aspect but rather through a complex interconnection of factors such as race, class, and gender, which disproportionately impact individuals from marginalized groups due to the intersection of these identities.

For instance, Crenshaw (1989) highlights that an algorithm trained on biased data can disproportionately affect not only women but also black women, who face compounded biases due to the intersection of their racial and gender identities. This intersectional discrimination, based on biased algorithms, can doubly exacerbate existing inequalities, affecting the hiring process and recruitment practices based on automated AI systems. By incorporating Intersectionality into the debate, we create a robust framework for understanding and raising awareness of the existence of multiple marginalized identities, which an algorithmic tool, particularly in employment and recruitment contexts, must consider avoiding exacerbating discrimination against certain demographic groups.

# 4. Methodology

In this section, we will address and explain the design and approach of our research, explain data collection, and also focus on the strategies used in this study, so that we can have a solid foundation for our ethical analysis of algorithmic discrimination in the context of employment and hiring practices.

Firstly, considering that we are conducting an ethical analysis, we opted for a qualitative research approach. This chosen methodological approach allows us to analyze and explore the complexities and differentiated realities of algorithmic decision-making processes, which can sometimes escape quantitative metrics. When focusing on fundamental aspects of societal life and individuals' daily experiences, which sometimes require a more subjective lens to encompass all individual peculiarities, it becomes critical that we analyze these themes—in this case, in the context of employment and hiring practices—not only through a technical perspective but also through a social, ethical, and legal lens. As Creswell (2013) stated, this approach is particularly valuable in understanding contextual factors and the sociotechnical dynamics that quantitative methods might overlook.

This qualitative methodology allows us an in-depth understanding to investigate the complex and multifaceted nature of the social and ethical concerns associated with algorithmic discrimination in automated decision-making systems through artificial intelligence, particularly in the context of employment and hiring practices.

In this chosen methodological approach, we primarily focused on data collection through real-world case studies of algorithmic discrimination in the context of employment and hiring practices. The meticulous selection of the case studies we will present enables us to conduct an in-depth and realistic analysis of specific cases where algorithmic decision-making systems led to discriminatory outcomes and perpetuated biases and algorithmic discrimination through their systems.

Using well-documented case studies of algorithmic bias from widely recognized companies such as Amazon, TaskRabbit, Fiverr, and HireVue brings individuals closer to a greater understanding of algorithmic systems, their mechanisms, and how they operate and manifest. The selection of these case studies was not only based on the potential of their documentation but also on the fact that they present different ways of reaching discriminatory outcomes (Barocas & Selbst, 2016) and, therefore, different ways in which algorithmic discrimination can manifest in various stages of the algorithmic decision-making process.

In the first case study, the automated hiring tool of Amazon, data collected from public media information and academic research articles, was identified as a mechanism of algorithmic discrimination because the tool penalized all resumes that included female gender, as its training and data collection had been based on resumes from male candidates (Dastin, 2018).

In the case of gig economy online platforms, TaskRabbit and Fiverr, data collected from platform policies, academic research articles, and public media information, both platforms were highlighted for how the algorithms present in these mechanisms were inadvertently favoring certain demographic groups over others. Specifically, how certain tasks and vacancies were being assigned to workers affected disadvantaged demographic groups (Rosenblat et al., 2017).

In the third case study on which we base our research, we present a different aspect of algorithmic discrimination present in these automated decision-making systems. Data collected from critical literature analyses, media, and public company information indicate that HireVue used artificial intelligence to analyze video interviews of candidates, including facial expressions, tone of voice, and characteristics related to ethnicity or race. This raised alarms and concerns about the existing bias in this system regarding cultural and linguistic differences that could be misinterpreted by artificial intelligence systems (Harwell, 2020).

In analyzing the data from the selected case studies, which were gathered through a systematic review of academic literature and publicly available information from these companies and the media, we compiled a comprehensive view of these case studies to provide a detailed examination of these algorithmic bias practices. In this research, we will use thematic analysis, as we believe this is the most suitable method for allowing a detailed interpretation of the collected data and enabling us to identify, analyze, and verify existing patterns. This facilitates the extraction of key themes of algorithmic bias present in our case studies and their respective ethical implications (Braun & Clarke, 2006). We find this approach quite appropriate, given that we aim to explore complex issues such as transparency, social justice, and ethical responsibility, which are central and fundamental to a better ethical analysis of artificial intelligence algorithmic systems.

Through thematic analysis focused on these real qualitative case studies, our strategy is to provide a robust and broad understanding that examines the complexities of algorithmic discrimination in the employment sector and goes further by offering insight into the systemic issues at play. This process facilitates the comparison of the different presented cases of algorithmic discrimination, highlighting the points of convergence and divergence, but which lead to the same result: a biased and prejudiced outcome. The main themes of this thematic analysis, as previously mentioned, are based on sensitive factors that discriminate against minority groups, such as gender and race, often inherent in the nature of prejudice itself and throughout other phases of the algorithmic process, which still carry discriminatory consequences for individuals, regardless of where they appear in the process, as we will see.

By focusing on real examples, we aim for the research to provide concrete evidence of how these automated mechanisms can perpetuate different forms of algorithmic discrimination. We intend to not only understand the causes and potential biases in AI-operated recruitment systems but also to emphasize the need to monitor the entire algorithmic process, from data collection, design, and implementation to understanding. Only in this way can we ethically intervene and mitigate any discriminatory outcomes. This approach will enable us to improve the performance of these types of selection and recruitment technologies in the employment context, understand the process, and enhance it. The central objective is for technological innovation to also promote justice, equity, and transparency and contribute to positive social change; if this does not happen, the involved parties should be held accountable in the use of artificial intelligence algorithmic systems.

# 5. Analysis and Findings

In this section, we will delve more deeply into our case studies and the central focus of our methodological approach for this research on algorithmic discrimination in the labor context and recruitment practices. We will present the results and highlight specific patterns in the real case studies on which we are basing our analysis, examining how algorithmic biases are present in them and their recruitment practices, how they manifest, how their mechanisms work, their flaws, and the respective ethical implications of algorithmic bias in decision-making systems and tools.

In this sense, as previously mentioned, we have chosen three case studies: Amazon's artificial intelligence recruitment tool (Dastin, 2018), gig economy platforms like TaskRabbit and Fiverr (Edelman, Luca & Svirsky, 2017), and HireVue. These cases have been meticulously selected to illustrate the different real facets of what could be discriminatory algorithmic processes, as well as the recurring themes and patterns highlighted in the specified case studies, thus providing a better understanding of the issue under study and the ethical and operational questions associated with algorithmic decision-making in employment and recruitment practices.

## 5.1 Case Study: Amazon

Firstly, we decided to present the case study of Amazon's artificial intelligence (AI) recruitment tool because we believe this meticulously chosen case is an excellent and illustrative example of how an algorithmic system can be highly discriminatory. Our analysis of this case, as mentioned earlier, was based on publicly available reports and academic articles. Amazon is a highly recognized company, as previously stated, and one of the major players in the technology sector, recognized as one of the "Big Five" largest and most important tech companies, alongside Google, Apple, Microsoft, and Facebook (Satariano, 2020).

Amazon used an automated recruitment tool based on artificial intelligence (AI) to streamline the hiring process, inherently increasing efficiency and reducing operational costs, employing it to review candidates' resumes. However, the problem arises when there is a bias in the training data, meaning that the data used to train the algorithmic system was based on resumes from previously selected male candidates[4]. The algorithm of this automated tool was trained to analyze resumes, predominantly from men, as they were the dataset the system had for training (Ajunwa & Greene, 2019).

Consequently, this artificial intelligence system learned to favor resumes belonging to men, considering them more qualified as they resembled the candidates chosen in the previous data and reflected historical patterns, excluding all resumes that included terms like "female." This systemic bias highlights the risk of reinforcing social stereotypes and perpetuating gender inequality in automated systems (Levy, 2018).

This represents a clear violation of the principles of justice, equality, and responsibility in labor practices, as stated by Binns & Gallo (2021).

We thus found that this is a great example of how the data and historical patterns learned by the algorithm can contain discriminatory elements and be perpetuated by the algorithm and in automated decision-making processes, especially in the context of employment and recruitment practices.

The case of Amazon's recruitment tool raises several critical issues. On one hand, it highlights the extreme importance of the data used to train any Artificial Intelligence (AI)

[4] The system has been trained for over 10 years based on male resumes (Ajunwa & Greene, 2019).

system, emphasizing that data must be meticulously selected to ensure it does not incorporate historical biases that produce discriminatory outcomes. On the other hand, this case also illustrates other significant flaws in these systems and algorithmic processes: the opacity and lack of transparency in automated decision-making processes. Amazon, in this case, lacked understanding and scrutiny regarding how decisions were made by the AI. There was, therefore, an extreme reliance on understanding the algorithmic decision-making process, and lastly, this case also underscores the urgent need for continuous monitoring and evaluation of all automated AI systems to detect and mitigate any existing bias in a timely manner, both in terms of data collection and the understanding and implementation of these systems (Raji et al., 2020).

As a company with a strong reputation, this Amazon case had implications for its reputation, as the negative publicity generated by the perceived bias in its algorithmic decision-making process led to public scrutiny, questioning the company's ethical and social mission, and also impacting consumer trust (Metz, 2018).

This case, as we observed, is quite rich at all levels and was meticulously selected for this ethical analysis. The Amazon tool case immediately raises ethical questions about the use of Artificial Intelligence (AI) in the employment sector and recruitment processes, particularly concerning social justice and ethical responsibility.

The system was systematically disadvantaging the same minority group, and the lack of transparency in this recruitment tool prevented both candidates and the company from understanding how decisions were made, raising concerns about the accountability of users of these technologies. This warning sign led the company to eventually abandon the use of the AI recruitment tool in its hiring and recruitment practices (O’Neil, 2016).

## 5.2 Case Study: HireVue

The second meticulously chosen real case study, in the employment and recruitment sector, that we present to enrich our ethical analysis of this investigation, is the case of HireVue's video interview platform. Our analysis for this case, as mentioned earlier, is based on publicly available reports and academic articles.

HireVue uses video interviews in its recruitment process, which are analyzed by Artificial Intelligence (AI) to determine if candidates are suitable for the selected position, based on a set of predefined criteria aimed at streamlining the recruitment process and reducing costs by automating the initial stages (HireVue, 2021).

However, this is another case that raises profound ethical concerns, as it has a tremendous potential to inadvertently perpetuate racial and cultural biases and their related assumptions. Again, if the algorithm's training data is based on a specific demographic group and does not consider candidates from underrepresented backgrounds, we may face algorithmic bias. Facial expressions, speech patterns, and even voice tone can have different cultural origins (Raghavan et al., 2020). We might interview candidates who are not native speakers or who exhibit speech patterns from different cultural backgrounds, which can lead to misinterpretations by the AI model, as it may not align with what is expected by the algorithm. Besides the potential for misinterpretation due to cultural differences, the lack of transparency in how algorithms evaluate candidates has been a major point of criticism (Binns, 2020).

This real case study of HireVue provides another perspective on how algorithmic discrimination can manifest if cultural and contextual factors are not taken into account during automated algorithmic processes. Algorithms that do not consider these cultural differences can produce results that are fully biased.

Ethically, this case also raises critical questions regarding the use of Artificial Intelligence (AI), particularly the need for AI systems to be sensitive to cultural factors and to be trained accurately and meticulously to account for behavioral patterns. Moreover, it highlights significant concerns about the privacy and consent of candidates regarding how their data will be used and how their characteristics are being analyzed (Kessler, 2018).

This case helps demonstrate the extreme necessity of ensuring transparency throughout the entire process to guarantee that stakeholders understand the automated recruitment process and its associated hiring practices, and that they provide informed consent for data processing through AI. It is equally important to reiterate the need for accountability among users of these automated decision-making systems, as what may be considered an innovative tool also carries social responsibilities in pursuing justice, equity, and transparency.

Similar to the previous case, and after recognizing the ethical implications this technology raised, it was discontinued in 2020, with the company emphasizing its commitment to transparency and fairness in all its technological tools (HireVue, 2021).

## 5.3. Case Study: TaskRabbit and Fiverr

The third meticulously chosen case study that we bring to the discussion of our research on algorithmic discrimination in the context of employment and hiring practices is the online freelance work platforms, namely Gig Economy platforms[5], TaskRabbit and Fiverr. We believe that this aspect of the freelance labor market also provides a case study

[5] The Gig Economy operates based on platforms that serve as an intermediary between the service provider and the end customer, usually are characterized by short-term, flexible and often precarious work. (Kalleberg & Vallas, 2018; De Stefano, 2016).

that reveals significant insights for our research.

The platforms, TaskRabbit and Fiverr, facilitate connections between clients and freelancers for a variety of desired services, managing task assignments and determining the visibility of freelancers on the platform (Edelman, Luca & Svirsky, 2017).

The analysis of this case study reveals crucial themes. These platforms not only rely on but truly depend on an algorithmic and automated decision-making system. Edelman (2014) and Van Doorn (2017), through their studies, shed light on how these platforms operate and essentially provide a deep understanding of the biases present in these automated systems. The algorithms on these platforms are designed to classify and recommend the most suitable workers for a given task, thereby influencing their visibility on the platforms based on their potential as measured by the algorithm. Both Fiverr and TaskRabbit use algorithms to make matches, yet the internal workings of these algorithms are quite opaque. Christo Wilson (2019) notes that workers from minority backgrounds, even with qualifications and ratings similar to those of majority group candidates, were less likely to secure certain job opportunities. As noted previously, such algorithmic bias may arise due to either the algorithm's design or the behavior of the platform users themselves.

This case once again underscores the importance of algorithmic transparency in automated decision-making processes, as otherwise, biases—whether inadvertent or intentional—have the opportunity to proliferate and exacerbate (Edelman, Luca & Svirsky, 2017). It is crucial for transparency to be part of all processes and to be accessible to users, through a clear process of how algorithmic decisions are made. On the other hand, we also see once again that users of these platforms play a fundamental role in their ethical responsibility not to reinforce existing inequalities through their own interactions and feedback processes (Hannák et al., 2017).

In analyzing these real cases of Gig Economy online platforms, such as TaskRabbit and Fiverr, we again delve into the concerns regarding the use of algorithms that disadvantage certain social groups, particularly minority groups. In these specific cases, the major ethical issues translate into the extent to which equity and justice are present on these platforms and how fundamental principles of algorithms are being violated and may perpetuate pre-existing social inequalities.

## 6. Mitigation Strategies and Recommendations

There is an urgent need to outline holistic and comprehensive strategies that can mitigate and even, in an ideal scenario, eliminate any algorithmic bias that may exist during any automated decision-making process based on algorithms trained by artificial intelligence, and in the context of our study, specifically within the employment sector and recruitment practices.

Pre-existing biases and those acquired throughout the training process of algorithms intrinsically affect society and fundamental aspects of individuals' existence. It is imperative that mitigation strategies are assertive and result in effective control of biases that threaten not only the efficiency and effectiveness of automated systems but also, consequently, the organizational structure for which these systems were designed, as well as the justice and social equity that should be represented in these automated algorithmic decision-making processes.

In this regard, this section will outline the mitigation strategies and recommendations based on this ethical analysis of algorithmic discrimination in the context of employment and recruitment practices, aiming for the ideal scenario of promoting justice, transparency, and accountability alongside the pace of the digital and technological revolution we are privileged to witness.

The first step in outlining the necessary strategies to mitigate algorithmic discrimination is to approach this process from a technical perspective and understand how we can introduce efficient and conscious notions of social justice into these automated processes, to ensure that algorithms do not inadvertently harm any demographic group.

Dwork et al. (2012) introduce the idea of "justice through awareness," meaning that to ensure algorithms can produce the desired equitable outcomes, we would adjust algorithms according to demographic parity and without any form of opportunity restriction. This aims to raise awareness of the algorithm and its design to balance justice and the most qualified possible choice within the most equitable context. We argue that an approach such as the one suggested by Dwork et al. (2012) would be a proactive step in combating algorithmic discrimination that can occur throughout automated decision-making processes that seek to ignore sensitive factors such as gender, race, age, etc., but which, as we have seen previously, often have the opposite effect of exacerbating pre-existing discrimination and reproducing biases.

Thus, the first of the mitigation strategies and recommendations we present is to examine algorithms and promote a conscious effort and explicit consideration of these sensitive attributes known to trigger social inequalities, particularly in our context, within the employment sector and automated recruitment processes. These attributes should be included in the design, conception, and implementation of any algorithm and artificial intelligence system used in automated recruitment processes to ensure the desired equitable outcomes.

The second mitigation strategy for algorithmic discrimination in the employment sector and recruitment practices is based on a fundamental principle that has been present throughout our ethical analysis: transparency.

Pursuing transparency in automated artificial intelligence processes is crucial for gaining trust in these processes and ensuring that all stakeholders feel integrated into a clear and transparent process. As Barocas & Selbst (2016) emphasized, it is essential for the involved parties to understand how algorithmic decisions are made and to recognize that if the process lacks transparency in all its phases—whether in the meticulously selected datasets, the selection of variables mentioned earlier, the training methodologies, and the design of algorithms, or in the understanding and implementation of these processes—there will be no clarity on how data is being trained and what the actual weights of our variables in automated decision-making are. This understanding helps us determine if we are facing algorithmic bias or if we are using an automated process fairly and equitably.

From one of the case studies presented earlier, the Amazon AI recruitment tool case, we observed that the algorithm had been trained in the same manner for over a decade, without recruiters realizing that the algorithm was reflecting biases based on the historical data selected for its training. Since most candidates for a long time were male, the algorithm assumed that gender, a sensitive attribute, was a fundamental characteristic in selecting the ideal candidate. Critically looking at this and any case involving technological processes, it is clear that technology is constantly evolving and transforming, resulting in technological innovations at sometimes astonishing speeds. Therefore, it is fundamental to advocate that all these technological processes, and in our case, automated recruitment processes using artificial intelligence, must be subject to mandatory, continuous monitoring that reflects the dynamic nature of technology and algorithms. This ensures that no one is excluded from the process based on factors unrelated to their actual qualifications for the job they are applying for. We also argue that these mandatory audits could, if not conducted by the respective entity, be performed by a third party, adding an additional layer of credibility and trust for a correct, impartial, and objective evaluation.

As we can see, the initial strategies for mitigating algorithmic discrimination in recruitment processes presented were technical strategies and recommendations. However, for an effective and robust fight against algorithmic bias in these processes, it is necessary to have a holistic and comprehensive response that includes solutions from various dimensions.

In the employment and recruitment sector, organizational policies also play a crucial role in reflecting the ideals that organizations aim to project as their guidelines. In this sense, all guidelines, especially ethical ones in this case, must be aligned with all parameters of the organization. Looking at automated decision-making processes within an organization, particularly in candidate recruitment systems, it is essential to adopt ethical guidelines that regulate the entire process of algorithm development and implementation in these systems, ensuring they align with the organization's ethical standards. It would be inconsistent for an organization that upholds human rights, social equality, and ethical responsibility not to consider how its trained algorithms for automated recruitment systems could potentially perpetuate social inequalities due to flaws in the algorithms. Therefore, it is important to emphasize that organizational policies must keep pace with their organizational practices.

In addition to organizational culture and ethical frameworks, it is fundamental that each organization not only communicates its values to its members but also provides them with the tools and conditions to contribute to reducing and mitigating the impact of algorithmic discrimination in the workplace. As Hanna et al. (2020) indicated, diversity within teams can lead to impartial outcomes in the development of automated system algorithms. We advocate that one of the strategies and recommendations for organizations would be to create diverse and inclusive teams. Such teams could lead to more robust results and more easily address and identify biases related to minorities, which are often overlooked in a fully homogeneous team from a majority demographic group. It is crucial to expand the sensitive characteristics of individuals within these teams to include a wider range of perspectives—social, ethical, demographic, legal, sociological, etc.—so that this representation within an organization helps address biases that others might ignore, thereby contributing to and not limiting social justice. We also recommend, from this perspective of providing tools and conditions to employees, that there be continuous and regular education and training programs for all employees. These programs should address algorithm development and automated artificial intelligence processes in the context of work, focusing not only on technical aspects but also on the ethical implications of disparities and inequalities that may arise from algorithmic bias.

We have discussed potential strategies and recommendations for mitigating algorithmic discrimination in the workplace, considering both the technical dimension of algorithmic systems and the ethical and organizational culture of those adopting automated artificial intelligence systems in the employment sector. Therefore, up to this point, we have looked at improving the experience of participants in this process.

Consequently, we also find it essential to consider the broader interventions that external regulatory bodies and governments should have in regulating and enforcing policies that protect algorithmic fairness. As Wachter et al. (2017) argue, the European Union provides a robust framework with the General Data Protection Regulation (GDPR), particularly concerning data protection. However, we recommend that similar regulations be developed in other dimensions to provide a strong legal foundation for combating algorithmic discrimination.

In the context we are analyzing—the employment sector—algorithmic decisions have a significant and profound impact on individuals' livelihoods, making it a highly sensitive area that should be closely scrutinized by regulatory bodies and governments. It would be extremely important not only to implement legal regulations but also to require organizations to submit detailed reports on the operation of their algorithmic systems and automated decision-making processes. In this way, governments and regulatory bodies could ensure that organizations are aware of the full nature of their AI-based decision-making processes and are vigilant against any potential algorithmic bias that may arise in the execution of automated models in employment. For such regulation to be fully effective, it should be accompanied by clear sanctions for companies that fail to comply with the legal requirements imposed by regulatory bodies. This would ensure that such policies are viewed as serious, assertive, and aware that society must address algorithmic discrimination as a daily issue impacting their lives, especially in contexts that challenge fundamental human principles. It would also hold companies socially accountable for using these technologies, promoting better practices that encourage companies to address and correct errors in these systems and improve them over time.

# Conclusion

Artificial Intelligence increasingly stands as a driver of innovation and profound transformations in the society we live in; today, it is a central and omnipresent element of a technological revolution that influences various aspects of daily life.

Our research, an ethical analysis of algorithmic discrimination in employment and hiring practices, has led us to explore algorithmic bias in these practices, at a time when the contribution of Artificial Intelligence to the current job market is undeniable, with companies seeking to streamline processes and reduce costs.

For this investigation, we initially relied on a theoretical framework based on two major theories: Critical Race Theory and Intersectionality Theory. This allowed us to obtain a robust and holistic understanding of the nature and causes of historical and systemic inequalities that can be perpetuated through seemingly neutral algorithmic decision-making processes.

By analyzing meticulously selected real case studies, not only due to the potential of their documentation but also because cases like Amazon, TaskRabbit, Fiverr, and HireVue present different manifestations of algorithmic discrimination (Barocas & Selbst, 2016), we found that despite companies increasingly aiming for more efficient, effective, and cost-effective processes, and trying to use technology to make more informed decisions, they may not be prepared for the profound ethical, social, and demographic implications that algorithmic decision-making systems can bring. These cases collectively demonstrate that algorithmic systems are not neutral tools but are intrinsically shaped during their design, notably through their training data and the contexts in which they are embedded (Noble, 2018). Despite the technological potential of these innovations, these systems can perpetuate and even exacerbate existing social biases, raising profound ethical and social concerns about their potential for biased results and, consequently, their impact on individuals' lives in such a fundamental aspect of human survival.

We found that for the main strategies and recommendations to mitigate algorithmic discrimination in employment and hiring practices to be effective, they need to be comprehensive and complementary, addressing technical, organizational, ethical, social, legal, and directive failures, and focusing on reinforcing the foundational concepts of any technological advancement, in this case, algorithmic: justice, transparency, and fairness.

We observed that it is crucial for the entire algorithmic process to be meticulously trained, with diverse and inclusive datasets to minimize any bias that might arise from skewed data samples, which negatively impact minority groups. It is also of utmost importance to foster transparency throughout the process for all stakeholders, providing clear and transparent information to help users not only better understand but also more effectively question the results of the algorithmic decision-making process, promoting crucial engagement with technological advancement and mitigating distrust in the process.

In addition to technical strategies for mitigating algorithmic discrimination, no single strategy alone will be sufficient. As previously mentioned, it is essential to develop a holistic approach to mitigation strategies. In this sense, one recommendation is also to create an inclusive (Binns, 2018) and diverse organizational culture and to provide training to employees on the ethical use of Artificial Intelligence. This is crucial to ensure a better understanding of these processes and, consequently, to mitigate the impacts of algorithmic discrimination. We also found that it is urgent to conduct regular audits and monitoring, especially in critical areas such as employment and hiring. It is essential to evaluate whether there are discrepancies and biases against any demographic group.

All the strategies and recommendations outlined throughout this research aim to achieve fairer, more transparent, and equitable outcomes in the context of employment and hiring processes. On one hand, they seek to leverage the technological potential of innovations, while on the other hand, they aim to protect against their shortcomings and improve them.

However, despite offering a critical perspective on the impact of algorithmic bias in the workplace and hiring practices, the study has some limitations. Notably, the selected case studies, while providing a robust and solid component, suggest that future research would benefit from expanding the number of cases analyzed. This would allow for a broader and comparative analysis of various sources and manifestations of algorithmic discrimination. Additionally, it would be interesting to conduct an impact study on the proposed mitigation strategies and recommendations to effectively assess their effectiveness.

This thesis thus contributes to the ongoing discussion about ethics in decision-making processes using artificial intelligence. Mitigating algorithmic discrimination in the context of employment and hiring practices is undoubtedly an extremely complex task and a constant challenge. The rapid evolution of technology leads to the emergence of new forms of algorithmic bias. Therefore, while the potential of algorithms to enhance recruitment in the employment sector—especially in terms of objectivity and efficiency—is significant, it is crucial that they are implemented by those who understand the social and ethical complexities involved. Continuous vigilance and monitoring of technology that profoundly impacts people's lives are essential.

# Bibliography


- Ananny, M., & Crawford, K. (2018). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *New Media & Society*, 20(3), 973-989. doi:10.1177/1461444816676645
- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org. Available at: https://fairmlbook.org.
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity.
- Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy." In *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, pp. 149-159. doi:10.1145/3287560.3287591
- Binns, R. (2020). "On the Apparent Conflict Between Individual and Group Fairness." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, pp. 514-524. doi:10.1145/3351095.3372864
- Buolamwini, J., & Gebru, T. (2018). "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR, 81, pp. 77-91. Available at: http://proceedings.mlr.press/v81/buolamwini18a.html
- Crenshaw, K. (1989). *Demarginalizing the Intersection of Race and Sex: A*

*Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*. University of Chicago Legal Forum.

- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/us-amazon-com-recruiting-insight-idUSKCN1MK08G
- Delgado, R., & Stefancic, J. (2017). *Critical Race Theory: An Introduction*. New York: NYU Press.
- Diakopoulos, N. (2016). "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures." *Digital Journalism*, 3(3), 398-415. doi:10.1080/21670811.2014.976411
- Diakopoulos, N. (2020). *Algorithmic Accountability: A Primer*. Data & Society Research Institute. Available at: https://datasociety.net/pubs/ia/DataAndSociety_Algorithmic_Accountability_Primer_2020.pdf
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). "Fairness through Awareness." In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. Association for Computing Machinery, New York, NY, USA, pp. 214-226. doi:10.1145/2090236.2090255
- Edelman, B., Luca, M., & Svirsky, D. (2017). *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*. American Economic Journal: Applied Economics, 9(2), 1-22.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ...

& Vayena, E. (2018). *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds and Machines, 28(4), 689-707.

- Friedman, B., & Nissenbaum, H. (1996). "Bias in Computer Systems." *ACM Transactions on Information Systems (TOIS)*, 14(3), 330-347. doi:10.1145/230538.230561
- Gebru, T., & Buolamwini, J. (2018). "The Coded Gaze: Bias in Artificial Intelligence." *MIT Media Lab Journal*. Available at: https://www.media.mit.edu/articles/the-coded-gaze/
- Gillespie, T. (2014). "The Relevance of Algorithms." In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-194). MIT Press.
- Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020). "Towards a Critical Race Methodology in Algorithmic Fairness." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (FAT* '20). Association for Computing Machinery, New York, NY, USA, pp. 501-512. doi:10.1145/3351095.3372826
- Hannák, A., Wagner, C., Garcia, D., Mislove, A., Strohmaier, M., & Wilson, C. (2017). "Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr." In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, New York, NY, USA, pp. 1914-1933. doi:10.1145/2998181.2998327
- HireVue. (2020). "HireVue Discontinues Facial Analysis Screening." *HireVue*

*Press Release*. Available at: https://www.hirevue.com/press/hirevue-discontinues-facial-analysis-screening

- Kalleberg, A. L., & Vallas, S. P. (2018). *Precarious Work and the Gig Economy: The Transformation of Employment Relations.* In S. P. Vallas & A. L. Kalleberg (Eds.), *The Handbook of Precarious Employment* (pp. 1-26). Routledge.
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
- Raji, I. D., & Buolamwini, J. (2019). "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products." In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, New York, NY, USA, pp. 429-435. doi:10.1145/3306618.3314244
- Rosenblat, A., & Stark, L. (2014). "Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers." *International Journal of Communication*, 10, 3758-3784. Available at: http://ijoc.org/index.php/ijoc/article/view/4892
- Satariano, A. (2020). *Amazon's AI Hiring Tool was Biased Against Women.* Bloomberg. Retrieved from https://www.bloomberg.com
- Sweeney, L. (2013). "Discrimination in Online Ad Delivery." *Communications of the ACM*, 56(5), 44-54. doi:10.1145/2460276.2460278
- West, S. M., Whittaker, M., & Crawford, K. (2019). "Discriminating Systems:

Gender, Race, and Power in AI." *AI Now Institute*. Available at:
https://ainowinstitute.org/discriminatingsystems.html