Fair design and deployment of machine learning algorithms on online labor platforms – What do managers do: A qualitative study

Author: Ellen Penkov
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

While the widespread adoption of algorithms on online labor platforms offers numerous benefits such as facilitating the efficient matchmaking of gig workers to clients, the potential harmful effects and uses such as bias-amplification and micro-management of workers remain relatively overlooked. Acknowledging the human influence at the core of each algorithm, this study seeks to gain an in-depth understanding of the individuals that are involved in decision-making surrounding algorithms on online labor platforms. By having conducted semi-structured interviews with platform managers, this exploratory research seeks to contribute to existing literature by providing insightful findings on how decisions are made, the motivations behind them as well as the practices used to ensure the fair design and deployment of algorithms on online labor platforms. This study revealed that managers are aware of algorithmic as well as societal biases on platforms and take various proactive approaches to mitigate the occurring biases. Moreover, it was found that despite the limited use of fairness tools as well as the lack of technical bias mitigation approaches in algorithm development, managers made a significant effort in ensuring platform fairness and showed general accountability for ensuring the fair design of algorithms. Overall, the findings highlight the complex nature of managing algorithmic fairness and underscore the need for further research and practical considerations in this area.

Graduation Committee members:
First Supervisor: Dr. J.A. Hüllmann
Second Supervisor: Dr. J.G. Meijerink

Keywords
Machine learning algorithms, online labor platforms, fairness, algorithmic bias, discrimination, decision-making

1. INTRODUCTION

Every aspect of our lives is being progressively influenced by machine learning (ML) algorithms: Making movie recommendations, suggesting personalized ads, the social media posts we read, or who to date (Mehrabi et al., 2019). Not only our private lives are affected but today, applications of machine learning can be found in nearly every industry (Jordan & Mitchell, 2015): from medicine and healthcare (Jiang et al., 2017) to banking and finance to transportation and retail (Richardson, 2021). Because algorithms take more factors into consideration than humans possibly can, machine learning is widely recognized for its capability to accelerate time-consuming processes, automate routine procedures, enhance the accuracy and efficiency of mundane tasks as well as assist in making better decisions. (Danziger et al., 2011). While machine learning algorithms can have a huge potential for good and are widely acknowledged for the elimination of human error, they are susceptible to bias and can only be as accurate as the data it is trained on. Historical biases, and insufficient or unrepresentative training data, for example, can lead to the amplification of biases in algorithms (Roselli et al., 2019). Because of this, algorithms can contribute to social injustice. In the past 10 years, there have been many instances of algorithms treating users unfairly and harming or discriminating against them based on personal characteristics, particularly race, and gender.

For instance, the COMPAS algorithm forecasts a defendant's risk to society and the chance of reoffending. It was found to predict higher risk values for black defendants than their actual risk (Angwin, Larson, Mattu, & Kirchner, 2016). In another instance, it was discovered that Google's Ads engine for targeted advertising served disproportionately fewer high-paying job ads to women than to men (Caliskan et al., 2017).

Just like in other domains, algorithms are being increasingly used on online labor platforms (OLPs) to optimize the efficiency of platform operations. Since the business model of OLPs relies on high-quality matchmaking and is crucial to the platform's performance, OLPs use algorithms to facilitate the efficient matching of supply and demand; in other words (gig) workers to clients (Möhlmann et al., 2021). This can mean matching, for instance, a driver to a customer on a ride-sharing platform such as Uber or suggesting the right worker to fulfill a customer's task on a freelance platform such as TaskRabbit. (Möhlmann et al., 2021; Park & Ryoo, 2023). Because of the algorithmic efficiency, more transactions take place, more consumers' needs are satisfied, more platform workers are paid, and consequently more revenue is generated by OLPs (Möhlmann et al., 2021; Stanford, 2017).

In addition to just matching consumers to service providers, platforms aim to control the behavior of platform workers by making use of algorithms; performance evaluations, behavioral cues, and customer reviews are commonly used to keep an eye on workers' activities and lead workers to exert behaviors that align with the company's goals. Besides algorithmic control, the already existing biases that are prevalent in society, in the form of race, gender, and social discrimination against specific groups is a serious problem that is not uncommon in traditional marketplaces and is repeatedly amplified by machine learning algorithms. (Möhlmann et al., 2021; Hannak et al., 2017). For instance, a study conducted on Airbnb found that accommodation applications from customers with African American names are about 16%-19% less likely to be approved than identical customers with distinctively white-sounding names (Edelman & Luca, 2014). Moreover, a study by Galperin and Greppi (2017) has shown that foreign job seekers are 42% less likely to win contracts from employers on the Spanish freelance platform Nubelo. Despite the increased use of algorithms, it is unknown to what extent online labor platforms account for the negative discriminatory as well as controlling effects of algorithms. Given the lack of behavioral research regarding the fair design and use of algorithms on online labor platforms and the decisions that preceded that, this study aims to answer the following research question:

What do managers do to ensure the fair design and deployment of machine learning algorithms on online labor platforms?

This qualitative research paper will add to our understanding of decision-making regarding the fair design and deployment of algorithms on online labor platforms. This will help address the current gap in the literature and provide real-world value to current research. While research has primarily focused on investigating the oppressive and controlling effects of algorithms on workers, as well as gender and social discrimination on online labor platforms, this research paper aims to understand the individuals involved in the decision-making process, their motivations behind these decisions, and the mechanisms they use in relation to the fair design and deployment of algorithms. This will be achieved through conducting semi-structured interviews with managers of online labor platforms, shedding light on the managerial practices, and providing valuable insights that contribute to the existing body of research.

2. THEORETICAL FRAMEWORK

In the following chapter, the theoretical framework will delve into the central question of what managers do to ensure the fair design and deployment of machine learning algorithms on online labor platforms by providing a coherent understanding of the dynamics between the interconnected concepts and conducting a thorough review using backward and forward snowballing. (Wohlin, 2014) The introduction of online labor platforms and the underlying machine learning algorithms is followed by an examination of fairness research and its various facets including fairness notions and algorithmic fairness. The various types of biases are then discussed, along with methods to reduce them, with an emphasis on software toolkits and checklist solutions.

2.1 Machine learning algorithms on online labor platforms

2.1.1 Online labor platforms

The emergence of online labor platforms in the past ten years has drastically altered how work is organized and accessed, having a profound effect on many facets of the labor market by challenging conventional employment models. (Berg, 2018).

Online labor platforms (OLPs), also referred to as digital labor or gig platforms, are multi-sided marketplaces that link on-demand, independent workers with clients or organizations that require their services. (Duggan et al., 2020; Meijerink et al., 2021). In other words, OLPs provide intermediation services between self-employed workers and organizations or consumers who wish to outsource fixed-term activities, for which gig workers receive monetary compensation. (Duggan et al., 2020; Meijerink et al., 2021). They include both location-based applications, which distribute labor to people in a given geographical region, and web-based platforms, where work is outsourced to a geographically dispersed population (Berg, 2018). Online labor marketplaces like Upwork, Freelancer, Fiverr, TaskRabbit, Deliveroo, and Uber are a few examples (Duggan et al., 2020).

These platforms, provide a variety of online services like writing, graphic design, and web development, or location-based services like ridesharing, food delivery, and home services (Meijerink et al., 2021). Whilst having many advantages for workers such as enabling more flexible work arrangements, there are also major emerging disadvantages such as job uncertainty, income instability, and the potential for algorithmic bias and discrimination. (Daskalova, 2018; Möhlmann et al., 2021; Jahanbakhsh et al., 2020; Monachou, 2019; Hannak et al., 2017).

2.1.2 The use of algorithms on OLPs

Just like in other domains, (machine learning) algorithms are being increasingly used on online labor platforms to optimize the efficiency of platform operations. (Möhlmann et al., 2021).

According to Oxford Languages, machine learning is "the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data". Since the business model of OLPs relies on high-quality matchmaking and is crucial to the platform's performance, OLPs use algorithms to facilitate the efficient matching of supply and demand (Möhlmann et al., 2021; Meijerink et al., 2021). This can mean matching, for instance, a driver to a customer on a ride-sharing platform such as Uber or suggesting the right worker to fulfill a customer's task on a freelance platform such as TaskRabbit (Möhlmann et al., 2021; Park & Ryoo, 2023; Stanford, 2017). Because of the algorithm, the more transactions take place, the more consumers' needs are satisfied, the more platform workers are paid and the more OLPs generate revenue (Möhlmann et al., 2021).

2.1.3 Matching & control

Platforms aim to govern the behavior of platform workers in addition to just matching customers with service providers. According to Möhlmann et al., 2021 algorithmic control refers to the use of algorithms to monitor platform workers' behavior and ensure its alignment with the platform organization's goals. Platforms attempt to regulate and keep an eye on most platform employees' activities through performance evaluations, behavioral cues, and customer reviews. Uber, for instance, makes use of algorithmic control by training their algorithms on behavioral drivers' data that consequently improves the algorithm's attempt to alter drivers' behavioral choices, which leads to more efficient, meaning more frequent, customer-driver allocation (Möhlmann et al., 2021).

While digital platforms allow workers to have more work autonomy, in the form of the flexible choice of tasks and work schedules they have the potential to gradually exert control over platform workers by making use of algorithmic management (Park & Ryoo, 2023). The so-called algorithmic management is used by OLPs to monitor and control the platform workforce. Möhlmann et al. 2021 refer to algorithm management as the "large-scale collection and use of data on a platform to develop and improve learning algorithms that carry out coordination and control functions traditionally performed by managers ". While tight algorithmic management has some advantages for platform organizations, such as maximizing worker efficiency and greater scalability, research indicates that it may also lead to workplace tensions in relation to compensation, workplace belonging, frustration among workers, as well as the feeling of lack of autonomy and flexibility to resist algorithmic instructions. (Park & Ryoo, 2023; Möhlmann et al., 2021).

2.1.4 Bias-amplifying effect of algorithms

OLPs increasingly rely on user-generated content, such as reviews to maintain quality control. Consider a situation where an algorithm gathers information about user interactions and decides how high on the search results gig offers are placed. Because of potentially biased user interactions and the positioning of results by algorithms, the top results gain popularity, and it remains unclear whether it is due to their inherent quality. biases (Spitko, 2019.; Olteanu et al., 2019). Workers with low ratings, which happen to be women and people from diverse backgrounds consequently, experience decreased visibility in top-search results, which negatively impacts their chances of being hired. (Hannak et al., 2017; Monachou, 2019), Outsourcing worker evaluation to consumers and using such input for automated decision-making creates a risk of reproducing user-generated biases (Spitko, 2019.; Olteanu et al., 2019). Despite the increased use of algorithms, it is unknown to what extent online labor platforms account for the bias-amplifying effects of algorithms.

2.2 Fairness

With the widespread use of algorithmic systems in our everyday lives, accounting for fairness has gained significant importance in designing and engineering such systems (Mehrabi et al., 2019). As humans become more dependent on and vulnerable to the decisions of machines, it is evident that biased or unfair algorithmic systems can have a systematic negative impact on society. To ensure the fair design and deployment of algorithmic systems, it is essential to first define what "fairness" means.

In a broader context, fairness is considered the "impartial and just treatment or behavior without favoritism or discrimination" (Oxford Languages, 2023). Fairness is not a clear-cut concept and has been challenging to define for scientists due to its multifaceted and context-dependent nature. (Feuerriegel et al., 2020) It is challenging to develop a single definition of fairness that is acceptable to all parties involved because varied attitudes and outlooks in various cultures favor different ways of viewing the concept depending on its context and domain. (Mehrabi et al., 2019)

2.2.1 Algorithmic fairness

By creating statistical definitions and algorithmic techniques to assess and mitigate biases, the field of algorithmic fairness, at its intersection of computer science, statistics, and mathematics, addresses the imperative to design algorithms that do not perpetuate discrimination, are free from bias, promote equitable treatment, and correct any inherent biases found in algorithms (Mehrabi et al. 2019).

Despite the difficulty of defining fairness, Mehrabi et al. (2019), who iterated and compared the 10 most widely used notions of fairness among researchers describe fairness as "[...]the absence of any prejudice or favoritism toward an individual or a group based on their inherent or acquired characteristics". Furthermore, they divided the proposed statistical notions of fairness into the following 3 categories.

(1) Individual fairness is based on the idea that similar people should be treated similarly. It highlights the notion that people with comparable traits or qualities ought to have similar experiences or outcomes. In other words, two people should be treated equally or similarly when decisions are being made about them if they are comparable in pertinent ways. Individual fairness is concerned with preventing unjustified distinctions between people and advancing justice on an individual basis. (Mehrabi et al., 2019)

(2) Group Fairness: Also referred to as demographic parity or statistical parity, group fairness focuses on guaranteeing fairness for various pre-established groupings within a population. It tries to prevent structural inequalities or biases in the decisions or results made for various groups based on protected characteristics like race, gender, or age. To ensure group fairness,

the proportions, or distributions of results among different groups must be comparable to or proportional to their population representation. For instance, if two demographic groups on an OLP have similar job qualifications, the likelihood of being hired by a client should be the same, in statistical terms. (Mehrabi et al. 2019; Kim & Cho, 2022a)

(3) Subgroup Fairness: In addition to individual and group fairness, there is also conditional fairness, commonly referred to as equalized odds or equal opportunity. It focuses on ensuring that a decision-making system's forecast accuracy is constant across several subgroups identified by protected traits. In other words, the predicted accuracy should be comparable across groups, showing that no subgroup is being unduly rewarded or penalized by the system (Mehrabi et al. 2019). In statistical terms, this means that no error type disproportionately affects any group. In other words, for members of the protected and unprotected group, the likelihood that someone in the positive class will be correctly assigned a positive outcome and that someone in the negative class would be erroneously assigned a good outcome should be equal. (Kim & Cho, 2022a)

2.3 Bias

In the subsequent section, a distinction between statistical and societal biases will be made. In order to emphasize the various types of biases that can occur across the 3 stages of algorithmic development, a framework that aims to classify biases depending on their stage of occurrence, along with a taxonomy of bias types is presented.

2.2.2 Statistical & Societal bias

As seen in the numerous examples above algorithmic fairness can be violated by biases, that can exist in many shapes in forms. Here, we distinguish between statistical and societal bias.

(1) Statistical bias is defined as "[…] a systematic deviation of an estimated parameter from the true value." (Feuerriegel et al., 2020). This is the case when the data continuously under or over-estimates the real value of the population parameter being evaluated. Statistical bias can be caused by several things, including errors in data collection, sample selection, or analytical methods. (Feuerriegel et al., 2020; Webster et al., 2022)

(2) Societal bias can be defined as "discrimination for, or against, a person or group, or a set of ideas or beliefs, in a way that is prejudicial or unfair." (Webster et al., 2022). That frequently depends on elements like race, ethnicity, gender, sexual orientation, socioeconomic status, religion, or other aspects of society. Institutional regulations, cultural norms, interpersonal interactions, and systemic disparities are just a few ways that societal prejudice can appear. (Webster et al., 2022)

2.2.3 Types of bias

In addition to the distinction of societal and statistical biases by Feuerriegel et al. (2020), Friedman and Nissenbaum (1996) proposed a framework to classify biases according to their occurrence in the technology development stage, including pre-existing, technical, and emergent biases.

(1) Pre-existing biases are societal and cultural biases that exist in the data and presumptions utilized to develop and operate algorithms. These biases, which might include systemic racism, sexism, or other forms of discrimination, can be the result of historical and cultural reasons and can be sustained by the data gathered and used to train algorithms. They exist independently and prior to the creation of the algorithm. Moreover, they are considered inherent and are a result of how people view the world.

(2) Technical bias refers to biases that result from the design and implementation of computer systems themselves as well as biases that stem from processing procedures and were added from the algorithm itself. The choice of characteristics, models, or training methods frequently introduces a bias that is not related to the practitioner but rather to the procedure's inadequacy to characterize the data. Such include limitations of computer tools such as hardware and software.

(3) Emerging bias describes circumstances where biases that were not caused by the data used to train the algorithm arise due to the use of the technology, typically sometime after a design is completed. This may occur for several reasons, including the introduction of new data that was not present when the model was first developed, unanticipated audiences, or changing social norms. To address emerging bias, technologies must be continually assessed and modified to guarantee that they are being used in an appropriate manner.

In the following figure, a taxonomy of the currently existing types of biases is depicted.

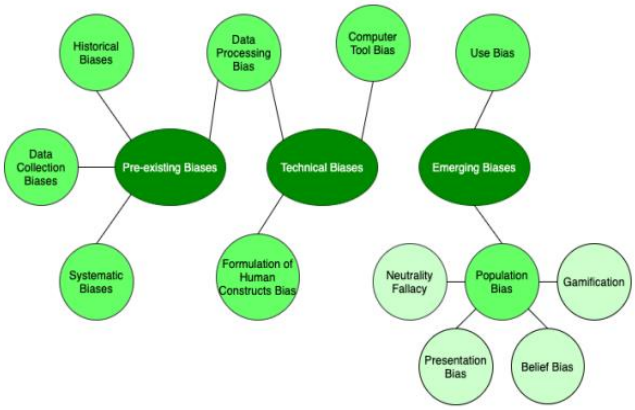

Figure 1. Types of biases (Richardson, 2021)

2.4 Bias mitigation approaches

In the following section algorithmic bias mitigation approaches are presented, followed by technical as well as non-technical tools for bias mitigation.

2.2.1 Algorithmic bias mitigation approaches

To address bias in algorithmic systems, a variety of techniques have been investigated in fairness research. These techniques can be divided into three main groups based on their application at various stages of the process: data collection, modeling, and output. These methods are commonly classified as pre-processing, in-processing, and post-processing approaches. (Kim & Cho, 2022a)

(1) Pre-processing approach: a learned model's erroneous performance can be directly attributed to specific characteristics of the training data. Pre-processing methods solve the problem by eliminating the bias present in the training data itself by modifying the data before it is used to train a model. This objective can be accomplished with a variety of approaches. These consist of data cleaning, data enrichment, resampling, and reweighting data rows, changing class labels for various groups, and excluding sensitive variables or proxies. (Kim & Cho, 2022a)

(2) In-processing approach: in-processing reduces the bias by adding a constraint to the learning algorithm or in other words, modifying the algorithm itself to make it account not only for accuracy but also for fairness. Techniques like modifying the decision threshold, adding regularization terms to the objective

function, or utilizing adversarial training can be used to achieve this. (Kim & Cho, 2022a).

(3) Post-processing approach: post-processing aims at only adjusting the outputs of a model, leaving the underlying classifier and data, in other words, the algorithm itself, untouched. Using Post-processing approaches gives developers the benefit of not retraining or remodeling an algorithm to ensure fairness. Supposing, a developed algorithm for task allocation on a ride-sharing platform is efficiently allocating rides to consumers, ensuring a fast match between supply and demand. Nonetheless, the development team notices that the algorithm favors male drivers over female drivers. With post-processing techniques, the team may adjust the outcome so that overall, the task allocation is more equal among males and females. Techniques, including, threshold adjustment, re-weighting, and calibration can be used. (Kim & Cho, 2022a)

2.2.2 Solution Space: Technical and non-technical fairness tools

Since the problem space of algorithmic bias is so large, a concentrated effort has gone into making fairness tools. Throughout the literature, various technical as well as non-technical solutions have been proposed by institutions and organizations that practitioners can use to embed fairness techniques during the design and deployment of (machine learning) algorithms (Richardson, 2021). These solutions come in two forms: software toolkits and checklists.

(1) Software toolkits serve as statistical and mathematical tools accessible via programming languages such as Python or websites that can be used to detect and/or mitigate biases throughout the machine learning (ML) pipeline. Commonly used toolkits are Fairness 360 by IBM, UChicago's Aequitas, and LinkedIn Fairness Toolkit (LIFT). (Richardson, 2021)

(2) Checklists are extensive guides created by fairness experts that developers can use to ensure the inclusion of ethical thought throughout the development and deployment of algorithmic systems. Some checklists are tailored specifically for data scientists or machine learning engineers, while others are intended for all parties involved in the project. Checklists involve questions and tasks to ensure that ethical considerations are considered throughout the project, from idea formulation to post-deployment auditing. (Richardson, 2021).

3. METHODOLOGY

3.1 Research Design

A semi-structured interview was designed to investigate what managers do to ensure the fair design and deployment of algorithms on online labor platforms. A semi-structured interview is a qualitative research method that combines a pre-determined set of open questions with the possibility for the interviewer to explore certain topics or responses further (Bishwakarma, 2017).

By enabling in-depth exploration of managers' perspectives, experiences, and practices related to fairness, the use of semi-structured interviews is a useful technique to ensure that crucial material is not omitted from a one-on-one interview while yet having a certain degree of flexibility. This allows the modification of (follow-up) questions to gain an in-depth understanding of what the company representatives think about algorithmic fairness and why (Bishwakarma, 2017).

In contrast to a strictly planned interview, a semi-structured interview's adaptable format enables one to inquire for more information or pursue a different line of inquiry that has been opened by what the interviewee is saying. (Fylan, 2005). The participant-centered approach inherent in semi-structured interviews yields data that is authentic and rich, which is crucial for a rigorous qualitative investigation, especially in the context of this study.

Additionally, semi-structured work well in delivering trustworthy, comparable, qualitative, and sensitive data from various participants (Fylan, 2005). That is essential for this study since participants might share not only their company's views on algorithmic fairness but also what they personally think is fair or not. Furthermore, all the data were gathered cross-sectionally to compare all findings at the same point in time.

3.2 Data Collection

3.2.1 Interview information

In total, 3 interviews were conducted, 2 of which were with representatives of 8vance and the remaining one with a company representative of Babysits. In the following sections, participants are going to be referred to as E1, E2, and E3. A detailed description of the interviewee's demographics and background is provided in Appendix B and screenshots of the platform interfaces are provided in Appendix C.

8vance is a Netherlands-based B2B workforce matching platform in the HR domain, whose primary business operations focus on efficiently matchmaking job seekers(talents) to various organizations (candidates) by using internally developed AI-driven technologies. By giving licenses to organizations, that sign up on the platform to seek talent, the platform generates revenue. Jobseekers, on the other hand, can sign up on the platform for free. 8vance provides services to a wide array of businesses, offering them the flexibility to search for talent both externally and within their own organization.

Babysits is a Netherlands-based, but globally operating childcare platform helping parents and babysitters. In other words, the platform is a match-making service that connects job seekers (babysitters) with job providers (parents). The company generates revenue by charging parents, that sign up on the platform, a subscription fee. Making a user profile on the platform is free for babysitters.

The interviews with 8Vance were conducted via Teams and lasted 1.02h with E1 and 1.15h with E2. The interview with E3 of Babysits lasted 45 minutes and was conducted on Google Meets. The interview protocol was sent to the participants beforehand. Because of that, there is a possibility of biasing interviewees' responses and making them less spontaneous due to having more time to prepare rehearsed or scripted answers. This might result in less insightful results. However, the protocol was provided on the interviewee's inquiry.

Moreover, the interviews were recorded and the parts corresponding to the theoretical framework were transcribed. The interviews were conducted in a group of 2 people from the HRM Bachelor thesis circle, and questions from all 4 members of the circle were asked to both company representatives of 8Vance. The interview conducted with Babysits was a one-on-one interview. Due to time constraints, only the questions relating to this research paper were asked to the company correspondent of Babysits.

3.2.2 Inclusion criteria

The participants chosen had to align with the following characteristics: An employee that is involved in either decision-making regarding algorithms (Manager) or is involved in algorithm development. (Developer)

Due to time constraints, the chosen sampling strategy was convenience sampling. The OLP employees were selected based on ease of access and those who could be interviewed as soon as possible. In total, 3 employees provided insightful responses for

this research. An overview of the participant's background is provided in Appendix B.

3.2.3 Interview Questions

In total 19 questions were asked to the interviewees (see Appendix A). At the beginning of the interview, 10 introductory questions were asked, aimed at gaining an understanding of the platform's business model, insights into the participant's background, position, experience, power in decision-making, tasks, and responsibilities as well as determining what algorithms are used on the platform. Next to the introductory questions, the remaining 9 questions were addressed to investigate this study's research topic.

The first 2 out of 9 questions were designed to work out the interviewee's experience with platform discrimination and what roles algorithms can play in amplifying that. Following that, question 3 was formulated to obtain an understanding of the interviewee's view on fairness in a personal as well as organizational context. Further, Questions 4 and 5 aimed at determining how the interviewee's fairness definition is embedded into the algorithms deployed and what is done during the algorithm development stage to mitigate possible biases. Furthermore, question 7 was posed to investigate when an algorithm is fair to use and how the manager makes sure certain fairness values are embedded into the system. Question 8 was dedicated to finding out what challenges designers face and lastly question 9 addressed the prioritization of fairness principles over developers' own careers.

3.3 Data Analysis

Deductive content analysis is a qualitative research technique in which the analysis of data, is guided by previously established concepts, also referred to as "top-down" approach (Vanover et al., 2021). Based on the theories developed in the theoretical framework, a coding scheme was created using Atlas.io, and during the analysis, it was applied methodically to the data to find patterns that fit into the predetermined categories. Initially, the predetermined categories were "Machine learning algorithms on OLPs", "Notions of Fairness", "Bias" and "Bias mitigation approaches", of which "Bias" and "Bias mitigation approaches" were further divided into the categories corresponding to the theoretical framework. Due to the emergence of new patterns in the data and the deviation of respondents' answers from the proposed framework, it was decided to implement a mix of inductive and deductive content analysis. Inductive content analysis describes a "bottom-up" approach. By identifying patterns that emerge from the data itself and establishing codes as analyzing the dataset that results in frameworks or categories afterward, inductive content analysis is applied. (Vanover et al., 2021). This approach allows for a balance between categories derived from existing theories and the emergence of new patterns from the data. (Proudfoot, 2022)

Corresponding to the original theoretical framework, "Machine Learning Algorithms on Online Labor Platforms (OLPs)" is the first section of the findings, laying out the algorithms used on the platforms and their various functions.

The findings proceed to examine the fairness notions expressed by the respondents. The data analysis showed that participants tended to advocate a broader, societal perspective on fairness as opposed to conceptualizing fairness in terms of algorithms. So that the categorization used accurately reflects the results, it is simply referred to as "Notions of Fairness" without further subdivision into the category of "Algorithmic Fairness."

In the bias section, respondents reported cases of bias that occurred on the platform in relation to algorithms. These were either societal biases or discrimination cases on the platform or a statistical or algorithmic bias as well as how they dealt with the stated cases. Hence why the findings were divided into "statistical bias" and "societal bias". The further categorization into the stage of the algorithm development they occur in as proposed by Friedman & Nissenbaum (1996) was rather difficult since participants reported quite specific and individual cases that cover multiple or no particular stages of the algorithm development and hence classifying the stated biases in "statistical bias" and "societal bias" provides more clarity for the reader.

The framework structure persisted, encompassing "Pre-processing", "In-processing", and "post-processing" bias mitigation approaches. The need for further classification into technical and non-technical practices, however, was justified by the interviewees' discussion of notable non-technical practices.

The pre-processing, In-processing, and post-processing approaches for algorithmic bias mitigation presented by Kim & Cho (2022a) are technical procedures. a section for "non-technical" approaches, depending on which stage they occur, was added to the findings. The reason for this is that participants tended to abstain from technical details and discussed rather general and platform-specific approaches around fairness and algorithms.

On top of that, it was investigated whether the participants used software toolkits, checklists, or other solutions. This section remains the same as proposed in the theoretical framework.

However, due to the flexible nature of semi-structured interviews, additional findings emerged beyond the predefined theory, leading to the identification and categorization of two significant aspects: decision-making and key challenges of individuals in relation to the fair design and deployment of algorithms.

4. FINDINGS

In the following section, the findings of the conducted interviews are presented. First, the various notions of fairness and the algorithms used on the platforms are outlined. Next, statistical, and societal biases are addressed, followed by pre-, in, and post-processing approaches aimed at mitigating those biases. Finally, additional findings including key challenges and decision-making around algorithms are presented.

4.1 (Machine learning) algorithms on OLPs

Both 8vance representatives explained the functionality of their internally developed core algorithm that provides skill suggestions based on CV information, like work experience, education, and skills, of the user. The user can decide to accept or reject the recommended skill on their profile, E1, and E2 elaborated.

Furthermore, E2 reported that the platform uses different kinds of algorithms, most of them are Natural Processing (NLP) algorithms due to the large amount of data that is textual. The NLP algorithm 8Vance developed was based on a pre-trained model and they added their "own data and techniques to fine-tune it", according to E2. Moreover, the platform focuses on large language models, trying to make its own variant of ChatGPT, specifically focusing on HR data. E1 highlighted that they are continuously working on (new) algorithms and are trying to improve their products.

E3 stated that the company developed a search and rank algorithm internally. The ranking is based on information, that the user provides on the platform, such as prior experience in childcare and education, as well as parents' reviews. The more information the user provides, the higher in the search they appear, E3 highlighted. In addition to that, E3 emphasized the

simplicity of the algorithm: „I mean, this is not some advanced algorithm but it's like a clear plus one plus one plus one when you completed this completed […] this is not magic. […] If you do that, it will give you this." He stated it is not a complicated AI model, but rather a simplistic algorithm.

In addition to the internally developed algorithm, E3 explained its platform utilizes Amazon Web Services (AWS) algorithms for authentication tools, which are mainly used for user verification. These algorithms are for" […] face recognition and working documents verification as well (as) for identity verification", E3 stated and clarified Amazon's service offers a variety of algorithms, that are used by numerous businesses.

Moreover, E3 mentioned: "We have one model we made ourselves now." E3 spoke of a model for document recognition that was created on their platform. Although the certainty level is currently quite low, this model is trained to determine whether a presented document is an identity card or not. According to E3, the model is still under development and is not live yet.

4.2 Notions of Fairness

According to the E1 fairness, especially in the context of the recruitment market, means algorithmic transparency. E1 thought that the existing hiring procedure is opaque and frequently relies on the intuition of recruiters without holding them accountable. If an algorithm comes to a result, it should be understandable to the user why and how the algorithm has made a particular decision, E1 explained.

In response to the question of what fairness means for him, E3 responded: "[…] Our platform is built around transparency, that's one of our core values. So, we try to build something which is transparent." He emphasized that everyone has the same chance of being ranked high in the search results and claims the platform is transparent on how this can be done. He explicitly stated: "That's fair to me. […]. Everybody can get as many jobs as possible."

E2 stated: "(fairness means) being impartial and equitable to everybody. […] everybody should be treated equally". Furthermore, E2 pointed out that quantifying algorithmic fairness is not an easy task because it depends on how well the algorithm complies with certain rules. He further explained that fairness is not solely dependent on technology but on understanding what fairness means in a broader societal context and incorporating that into the system is the real challenge.

4.3 Bias

4.3.1 Societal bias

The interviewee emphasized that although biases may exist in the data, they are not actively promoting gender discrimination and no overt gender-based discrimination has occurred on their platform. There is no instance, that E1 knew of that a man was preferred over a woman by a recruiter, or the other way around. He further stressed, that there are no "complaints or a case where a choice was made based on reasons that are […] unfair" and attributed that to the fact that no personal information, such as gender, ethnicity, race, or religion is displayed on the user's profiles.

E3 reported an incident in which parents rejected a male babysitter from the Netherlands because they preferred a female babysitter. E3 admitted the male babysitter's communication style may have been one of the reasons for the rejection, though. "So yeah, that's, in that regard, discriminating babysitters", he stated and elaborated:" Lower reviews are given to men, at least, according to our knowledge, this happens. If you're a woman, you're more likely to get a babysitting job. […] You will have more bookings and more reviews. And if you're a man, it will be harder to get started to get bookings and reviews. You will be lower on the search results." He finally concluded: "That's not something we will do something about." E3 stated what the babysitting platform did in response to the gender disparity on the platform: "We removed the attributes of gender from all profiles." Initially, the platform required users to provide information about their gender but E3 explained "[…] just from the profile photo, people are able to see the gender of a person". Nonetheless, the platform decided to remove that attribute to prevent discussion around that topic.

4.3.2 Statistical bias

E1 mentioned that men and women tend to define their profiles differently, with men often including skills they do not possess, which might create a bias in the data that one cannot account for.

In another instance, E2 reported that when training a model for a nurse case vacancy, the model consistently favored a woman's profile over a man's one. He found that the model's preference for female profiles can be attributed to the training set of data. He clarified that the job postings from LinkedIn were used to train the model, and it turned out that most of them mentioned women. Because of that bias, the project was eventually, discarded due to the unfairness aspect, E2 reported.

Moreover, E3 expressed doubt regarding the face recognition algorithm's capacity to determine whether someone has their eyes open. He mentioned instances in which he could clearly see that a person's eyes were open in a picture, but the algorithm misidentified it. E3 did not consider document recognition or image recognition to be discriminatory since it simply analyzes whether a face can be recognized or not. He explained that the algorithm is not always 100% accurate and said, "that's not discriminating". Users get a notification in case the photo is rejected, and they have the option to upload a new picture, he described.

Moreover, E1 stated that they cannot impact the validity of the user's input on their platforms. He elaborated that LinkedIn, for instance, requires some type of proof of education or degrees from users, but on their platform, they simply trust the jobseekers with what they claim to know or the skills they have. He admitted that this could create some bias, and there is nothing the platform currently does about it.

4.4 Pre-processing

4.4.1 Technical

The company representative explained that during the data collection process, they need to make sure that the population is represented in the sample to ensure an, as far as that is possible, unbiased algorithm. E1 explained: "(…) in the data collection, we try to always gather as much data and also as much diverse data as possible." Furthermore, he underlined: "That's actually the most important because the data in a large part determine the outcome of the algorithm. So, the algorithm itself is not inherently biased or unfair […] it just learns." "[…] We say if you put garbage in you also get garbage out.", E1 exemplified. E2 shared the same opinion about the significance of identifying and mitigating bias while acknowledging its presence in all data.

Furthermore, in the pre-processing stage, the recruitment platform "remove(s) any kind of personal information", which are sensitive attributes, like gender, race, ethnicity, or religion, as well as the date of birth of candidates to prevent discrimination, both interviewees pointed out. In another step, developers clean the gathered data by "removing unwanted characters", E2 elaborated and further elucidated that this step is essential to avoid a biased algorithm.

4.4.2 Non-technical

The algorithmic decisions are based on CV data from talents, like skills and work experience E1 emphasized and said that this is "information that of course is discriminatory, but you also have to discriminate on that in order to find the right candidate." E1 explained they cannot influence what information users put on their CVs or what skills they claim to have. The validity of the data is something they cannot influence, as E1 and E2 explained. They both stated they rely on honest user inputs.

E3 expressed that the data they use as input for the search and rank algorithm, cannot be fairer than it is now. He also stated, they rely on honest user inputs and claimed that it is a straightforward process: "The more information a user provides the higher in the ranking they appear. "

4.5 In-processing

4.5.1 Technical

To further ensure the accuracy of the matchmaking E1 explained how they consolidate synonymous terms under a single category to facilitate matching: „So even if the company uses skill term A and talent uses term B […], we still find a match". He also stressed that there is ongoing maintenance to "keep it (the algorithm) up to date with the real world" and further emphasized that "that's an intensive process". The maintenance of the algorithm is done manually by the team, based on real-world data from vacancies and talent profiles.

E2 stated that the procedures they are applying like removing sensitive attributes and data cleaning are good, but they are trying to make the data pipeline more robust. E2 also pointed out that the data always change; therefore, the processing techniques also must be adjusted constantly.

4.5.2 Non-technical

Since E3 stated fairness to him means platform and algorithmic transparency the interviewer further inquired what the platform discloses and what not. E3's responded: "We should disclose more […] we don't have a dedicated text about that on our website". He further stated what they disclose is how a user becomes a "super sitter", which eventually means that they can get on top of search results. "Yeah, so to me fairness is that that we communicate to every user if you complete all your badges. And if you want to become a super sitter, this is what you need to do. And that's clearly explained. ", E3 stressed.

E3 explained that the profiles should include details like age, profile picture, general description, education level, and experience in childcare, preferably verified identity, and criminal records documents in order to appear high in search results. Additionally, he stated, active users, as well as users who fill in non-mandatory information, such as a profile description, or provide a video, will increase their chances of appearing on top search results. There are also some criteria, that will not influence the ranking algorithm, such as the experience with children with special needs.

4.6 Post-processing

4.6.1 Technical

After the algorithm has been developed the team of developers applies the train-test-split which is a "standard practice in training AI models" to evaluate the machine learning algorithm, E1 stated. E1 explained they "train the model on a subset of […] usually 70% of the data". The remaining 30% of the data is kept separate for evaluation of the model and further assesses the model's performance and accuracy. Furthermore, E1 and E2 indicated that the post-processing approach, hence the evaluation metrics, is dependent on the person developing the algorithm.

E2 admitted: "We don't have a very extensive process of checking out whether our AI is really working or not, because we don't have that many resources. But we do a preliminary check whether it is working as expected or not."

E3 reported that they use an algorithm to assess profile pictures on their platform. He mentioned that "we don't want to allow profile photo with some violence. […] or explicit content". If the algorithm detects such content and assigns a value of 95 instead of 99, the platform reviews it manually to determine the appropriate action.

Moreover, E3 expressed doubt regarding the face recognition algorithm's capacity to determine whether someone has their eyes open. If the accuracy is not 100%, the algorithm rejects the profile picture, and the data is checked manually by employees.

Moreover, E3 mentioned a unit test, that they do for their search and rank algorithm. He stated: "If you put in this data, it will result in this score." He explained the score should be the same given the same data input. If that is provided, the algorithm is fair, according to E3.

4.6.2 Non-technical

By providing a ranked list of match results based on objective criteria like overlapping skills and relevant work experience which can always be adjusted by the user, E1 and E2 hope to uphold fairness. E1 clarified that the user can either accept or reject a skill, that was recommended by an algorithm: "They are making decisions for themselves, we are just trying to present the most relevant information, which they can select from, but they always have an option to really select or discard the options altogether […] because we always give them that option on the screen", E2 explained. E1 stressed the importance of involving the user in the decision-making process. He conveyed that their approach involved seeking "extra verification from the user" when making suggestions based on algorithmic recommendations. He highlighted the commitment of the company to maintain user control by "always try[ing]to keep the human in the loop". Furthermore, E1 stated: "Users can also provide feedback on match results, indicating whether it's a good or bad match." The platform aims to incorporate feedback into the algorithms. The company makes sure to distinguish between objective match quality and individual preferences, recording the latter separately to avoid incorporating it into the algorithm.

If a recruiter decides to not choose the top candidate provided by the algorithm but decides to hire the 2nd or 3rd one, the platform can hold the recruiter, that is also a customer on the platform, accountable for their choice, E1 explained. This has not happened yet, but due to the transparency of the algorithm, they have the possibility to encourage recruiters to explain their choices and justify their decision if a discriminatory case might happen, E1 stated. However, the interviewee acknowledges that the decision to choose a candidate rest with the recruiter.

4.7 Software Toolkits, Checklists & other solutions

E2 mentioned that he had heard of a fairness, accountability, and transparency framework, but adhering to such is not binding for the company, but rather aims as a general recommendation. E2 acknowledged the existence of frameworks and expressed that he wished to incorporate those, but it is not an easy task. E2 said they follow what is generally advised in fairness research and they try to incorporate that, but there is nothing concrete they do. E2 described the incorporation of such frameworks as rather flexible and expressed the wish to make the data pipeline more robust. Furthermore, E2 added their team makes use of open-source frameworks to check for potential biases that might have been produced which "have been tested by many people around

the world. […]". "And we rely on that", E2 asserted. E1 stated: "We don't have an external framework […] we kind of do that ourselves."

E3 clarified: „We don't have a guideline document. The leadership decides(that). We're not that professional yet.", and in addition to that he stated that their team makes decisions on what is (un)fair and that they do not have any type of framework they adhere to.

4.8 Additional findings

4.8.1 Key Challenges

E1 elucidated that one of the key challenges is getting enough significant data. He further clarified that it is "difficult to gather data that is representative". Despite having millions of profiles in their database, evaluating algorithms based on user feedback is difficult because it relies on a small number of users.

E2 found it hard to define fairness in the context of algorithms since "Every system comes with a problem in general. Because room for improvement is always there". E2 pointed out that quantifying fairness is not an easy task because it depends on how well the algorithm complies with certain rules. He further explained that fairness is not solely dependent on technology but on understanding what fairness means in a broader societal context and incorporating that into the system is the real challenge. He further explained that if "90% of the time, it(the algorithm) is good,10% it is not good" the team cannot deploy an algorithm. He said: "This can pull us back", implying that this can slow down their progress.

4.8.2 Decision-Making

In response to the question of who decided whether an algorithm is fair to use, E2 declared that whether to use an algorithm or not remains a human management decision. He made it clear that while human judgment plays a role in the decision-making process, they nevertheless hope to influence and direct that choice through metrics and performance assessments of the algorithm. He also claimed that the choice is consistent with the company's mission and aims, which center on increasing transparency to various stakeholders. E1 conveyed that the responsibility for making the decision to use an algorithm lies within the organization, to be precise it is handed internally, either by E1 himself or their team. He further stated that "[…] (they)don't use an external framework." To further validate whether the algorithm is fair to use the team relies on their own judgment by evaluating whether what they see in the predicted outcome of the algorithm aligns with their own observations. The interviewee stated that the "research team set criteria" to assess the efficiency of the algorithm. E2 explained that decisions regarding algorithm design and metric choice are made by a team of six and to do so regular meetings are held. "People generally presented results in front of our team[…] and we take a cumulative decision", E2 explained and further emphasized they do not "go towards unethical things". "I'm kind of proud we have a diverse team", E1 stated. Both interviewees of the recruitment platform highlight the diversity and international makeup of the team and emphasized the beneficial effect it has on decision-making.

E3 expressed a similar strategy: "We don't do other (things) than that, I mean, we're a small team of 20 people with a leadership team of three, four people." What is fair or not and which decisions to make around algorithms are eventually decided by discussing certain topics with a team, E3 said.

5. DISCUSSION AND IMPLICATIONS

This study's purpose was to explore what decisions managers do surrounding the fair design and deployment of algorithms. In the following section, the results of the conducted interviews will be interpreted by thoroughly discussing and systematically comparing the results with existing literature. By examining the insights obtained from the discussion, the broader significance and applicability of the study's findings in both theoretical and practical contexts is presented.

5.1 (Machine learning) algorithms on online labor platforms

The research findings showed that online labor platforms use algorithms for a variety of purposes. Algorithms are used for document authentication, ranking, and search, skill suggestions based on user-provided information, and most notably for efficient matchmaking processes. The algorithm used on the platforms is tailored to the specific functions and objectives of each individual platform. The findings underscored the critical role algorithms play in platform operations, particularly in increasing platform operation efficiency and facilitating the process of matching supply and demand. This finding strongly aligns with the theory expounded by Möhlmann et al. (2021), which accentuates the rising importance of algorithms for platform performance and operations.

Moreover, the data yielded a rather unanticipated outcome. While the technology-driven recruitment platform developed its algorithms internally, while also demonstrating a higher level of complexity, the babysitting platform, in contrast, combined internal algorithm development with utilizing external algorithmic services from 3rd party providers, such as Amazon. The internally developed search and rank algorithm by the babysitting platform exhibited a notable simplicity in its design.

The data suggested that the decision to develop algorithms internally or outsource their services depends on the complexity of the services provided and the company's overall objectives and value proposition.

An additional explanation for outsourcing can be attributed to the technical expertise of the platform. Outsourcing allows OLPs to leverage the quality and knowledge offered by external service providers. There is simply put no need to develop an algorithm internally if an external service provider, already does so offering a high-quality and cost-effective service.

Möhlmann et al., 2021 investigated the algorithm used on the well-known ride-sharing platform Uber. Their findings show that algorithms are used to micromanage and exert control over gig workers. The results might suggest that the babysitting platform incentivizes or leads users for completing their profiles, by providing them with a "badge" to become a "super sitter". This leads to higher visibility on the platform's search results, which leads to more frequent matches, which eventually leads to more revenue for the company, which is how OLPs commonly generate revenue, according to Stanford (2017). Nonetheless, it is important to view this case critically since the data from this study does not offer proof for this claim. The platform representative claims that they are completely open and transparent about how users can appear more prominently in search results. Although they don't have a dedicated section on their platform that is specifically for this purpose, they are willing to give users more information about the algorithm's decision-making process if they inquire. The platform emphasizes that there is a greater chance for fair outcomes the more information and experience a user provides, along with positive reviews from parents. Because the platform shares information about the algorithms—which are, in fact, simplistic in nature—openly, this cannot be characterized as micromanagement or control of gig workers.

5.2 Notions of Fairness

The study's findings showed that managers' opinions on fairness in online labor markets varied widely. Two participants heavily emphasized the value of algorithmic and platform transparency, while one participant offered a broad and socially centered understanding of fairness. The latter assimilates the definition of individual fairness by Mehrabi et al. (2019). In addition to that, one manager found it hard to define fairness in the context of algorithms, since the goal and objectives of each algorithm differ depending on the fairness rules used to train it. This aligns with the findings of Mehrabi et al. (2019), who listed 10 different statistical definitions of fairness and emphasized that the choice of which one to choose is difficult and remains with the practitioner. Overall, the definitions of fairness by managers diverged from the theoretical framework because they were rather focused on social and platform-centric factors than on technical, mathematical, and statistical ones as suggested by theory. (Mehrabi et al., 2019; Kim & Cho, 2022a). According to Feuerriegel (2020) and Mehrabi (2019), the findings, that show the varying views on fairness, are consistent with the idea that fairness is a multifaceted and context-dependent concept that varies across individuals.

5.3. Statistical & Societal Bias

5.1.1 Societal Bias

Both online labor platforms claimed they do not actively promote gender discrimination. Nonetheless, the results showed that one of the 2 examined OLPs has gender disparities. On the babysitting platform, women tended to receive higher reviews from parents and consequently get more jobs since they appear higher in search results. The observed preference for women on the babysitting platform serves as an illustration of the bias-amplifying effect of algorithms. As stated in the theoretical framework, platforms heavily rely on user-generated content, particularly reviews, which have a direct impact on how visible and highly ranked gig posts appear in search results. (Spitko, 2019; Olteanu et al., 2019). This study's finding contrasts with earlier research by Hannak et al. (2017) and Jahanbaksh et al. (2018), who found higher review ratings for men on well-known platforms, which is noteworthy. According to Chaudhuri and Gangadharan (2007) women are more trustworthy than men, hence why a possible interpretation of this study's finding might be the greater trust platform users have for women, particularly in the context of childcare responsibilities.

A significant observation, however, is the Babysits managers' lack of concern about the gender disparity on the platform. This phenomenon could be attributed to the notion that the algorithms merely mirror the existing societal attitudes (Friedman & Nissenbaum, 1996) and therefore managers do not perceive that as an inherent unfair issue. Generally, society perceives women to be more disadvantaged than men and since gender disparity occurs the other way around, it might not be viewed as an issue. In response to the bias-amplifying effect of algorithms, platforms could possibly interfere with manipulating algorithmic processes or implementing fairness-oriented classifiers, which could be a possible in-processing approach to account for fairness as suggested by Kim & Cho (2022a). Nonetheless, this is not the case.

What managers did, however, is take action to address the issue by removing the characteristic of "gender" from their platforms Despite the possibility that users can deduce a user's gender from their profile picture, both platforms have chosen to remove this feature to promote equality and lessen any difficulties or biases associated with it. The managers of the platforms hope to create a setting where people are assessed based on their skills and qualifications rather than their gender. This reflects a proactive approach to addressing gender bias and may be attributed to the current debate and discussion surrounding gender.

5.1.2 Statistical Bias

The results showed that the platforms take proactive steps to address statistical bias: Managers acknowledge instances in which statistical bias was found in their algorithms, such as the occasional inaccuracy of the face recognition algorithm on the babysitting platform, or a model's preference for female profiles on the recruitment platform. In response, the recruitment platform stopped using the algorithmic model altogether by scraping the project, while the babysitting platform compensates for the algorithmic bias by checking certain instances manually.

On the other hand, there is a statistical bias they did not actively do something about. While OLPs have control over the validity of some inputs such as identification documents on the babysitting platform, there are other user inputs such as CV information on the recruitment platform, where they rely on honest user inputs. This consequently introduces statistical bias and algorithmic inaccuracy since the algorithm is fed with data, that is not 100% reliable. (Webster et al., 2022). The data suggested that managers are aware of that, but instead of seeing the need to address this potential bias, they continue to place their trust in honest user inputs. The recruitment platform did not express a desire to implement similar measures, despite being aware that platforms like LinkedIn oblige users to verify CV information.

5.4. Data Processing

In the following section, it is summarized what managers do in each stage of design and deployment of algorithms to ensure fairness, combining technical as well as non-technical approaches. Finally, the common pattern among these findings is discussed.

5.1.3 Pre-Processing

The technology-driven recruitment platform emphasized the significance of a large and diverse dataset for the training data, as well as the removal of unwanted characters and sensitive attributes, such as race, gender, ethnicity, and religion during the pre-processing stage to ensure the fair design of algorithms. The babysitting platform does not pre-process its data for the internally developed algorithm. They use the data given by the user as direct input for the algorithm.

5.1.4 In -Processing

The recruitment platform shows a proactive approach to ensuring algorithmic fairness by regularly updating its algorithms to account for changes in the real world as well as adjusting the processing techniques. Additionally, the recruitment platform places a strong emphasis on incorporating user feedback into the development of algorithms.

The babysitting platform, in contrast, places a greater emphasis on ensuring transparency to users. It is important to note, however, that this platform does not formally disclose any information regarding transparency on its website. The babysitting platform shows a strong commitment to transparency by having the ability to directly explain its algorithm upon request, even though no customers have made this request thus far. Although the platform does not currently have a section specifically devoted to "transparency" on its platform, it acknowledges the need for improvement in this area and expresses the idea of adding one. This observation may be explained by users' lack of requests for information surrounding transparency. These results demonstrate the various approaches taken by the platforms to encourage fairness in algorithmic development.

5.1.5 Post-Processing

The recruitment platform uses a train-test-split approach to assess the performance of its machine-learning algorithms. At the same time, it acknowledges that no perfect performance can be completely guaranteed. The babysitting platform does not use direct testing because of the simplicity of its internally developed algorithm. They do, however, step in and manually check the accuracy of face recognition or document accuracy if the outsourced algorithm does not achieve a perfect accuracy score of 100.

Additionally, the recruitment platform adopts a user-centric approach by making algorithmic skill suggestions rather than fully assigning them to job seekers. Users can exercise control and make informed decisions thanks to this strategy, which places an emphasis on keeping humans in the loop. On the other side, the transparency of the algorithm also makes it possible for recruiters to be held accountable if they choose not to hire the suggested candidates. Even though the recruitment platform does not exercise that, they could possibly do so if a user experienced discrimination. The recruitment views involve the user in the algorithm development as well as ensuring transparency as a key component for platform fairness.

5.1.6 Data Processing: Combined discussion

Combining the results of bias mitigation approaches of the three stages of algorithm development, the following conclusions arise from the data.

Across both companies, there was generally little discussion of technical pre-, in-, and post-processing steps. However, it is noteworthy that throughout all three stages, the technology-driven company showed a greater emphasis on technical procedures. The babysitting platform, on the other hand, made no mention of the technical data processing or design methods it used for its algorithms. There are two possible causes for that.

First off, the recruitment platform prioritizes technology more and creates its algorithms in-house. These algorithms tend to be more complex since the training data is mostly textual. The babysitting platform, in contrast, uses a simplistic algorithm, that does not require the processing of complex data. Given the simplicity of the algorithm, there is no need for bias mitigation approaches since the probability of bias is low.

Secondly, the 2 company representatives of the recruitment platform, are part of the engineering team and generally had more technical expertise. The interviewee from the babysitting platform is the CEO & Founder and likely focuses on platform operations not directly related to algorithm development.

Overall, the results indicate that, despite the limited use of technical approaches in algorithm development, managers consistently made a significant effort to ensure platform fairness and no evidence of the use of biased or unfair algorithms could be found.

5.4. Software Toolkits, Checklists & other solutions

The results showed that although open-source fairness tools have been the subject of extensive research, particularly in the context of technical solutions (Richardson, 2021), the OLP managers interviewed showed awareness of these (open-source) tools but do not actively use them to ensure fair algorithms as they did not perceive them as binding requirements.

While the management of the technology-driven recruitment platform views frameworks as general suggestions, expresses a desire to include fairness frameworks, and recognizes their value, they encounter difficulties in incorporating them into their algorithms. The data showed that the platform uses open-source frameworks partially. However, which ones specifically they use is not clear. The results also suggested that rather than rigidly adhering to a single predefined fairness framework, both platforms make use of their own knowledge and the talents within their teams. Hence why one might conclude, the managers' reliance on their own abilities and unwillingness to rely on outside frameworks may also be factored into why they don't use these tools.

Richardson's (2021) thorough literature review, which sheds light on the difficulty faced by developers when faced with a wide range of fairness tools, might be an explanation of the findings of this paper. His work highlights how practitioners frequently feel as though there are too many options available to them.

5.2 Additional findings

Despite not directly answering the research questions, key challenges and decision-making surrounding algorithmic fairness are discussed here due to the explorative approach of this study.

5.2.1 Key challenges

The data showed that, managers of online labor platforms acknowledged the challenges associated with gathering representative data and conducting effective evaluations of user feedback. Moreover, they highlighted the difficulty in defining fairness, particularly within the broader societal context, and emphasized the difficulties of incorporating fairness into machine learning algorithms. The managers recognized that even a minor error in the algorithm could significantly slow down progress in algorithm development.

5.2.2 Decision Making

The findings indicated that, despite the use of metrics and performance evaluations by the recruitment platform, human management is still responsible for deciding whether an algorithm is fair to use. The data also suggested that without using external frameworks, the team in charge of algorithm design and deployment establishes fairness standards, considering OLPs missions and goals also as a guide for decision-making. These democratic procedures guarantee inclusive decision-making that considers the knowledge and values of the team.

5.3 Implications

5.3.1 Theoretical Implications

This explorative study expanded the academic knowledge of what is done by managers to ensure the fair design and deployment of algorithms on online labor platforms.

The findings of this paper contribute to the literature of Möhlmann et al. (2021), who investigated the oppressing effects of algorithms on gig workers. This study adds to that, by investigating what decisions precede algorithms, that potentially can harm people.

Moreover, this paper significantly adds to the extensive literature review on fair AI solutions conducted by Richardson (2021). Richardson's work examines a wide range of technical solution tools, practices managers can do to ensure fair AI, as well as the challenges developers face when using the suggested fairness tools. This paper deepens the understanding by shedding light on managers' actual practices surrounding fair algorithms.

As academic research in this field is limited, this study is a starting point for further investigation into the decision-making practices surrounding fair algorithms on online labor platforms.

5.3.2 Practical Implications

The study's findings highlight managers' general accountability for ensuring algorithmic system fairness. Nonetheless, the potential for further improvement exists.

It has been found that despite their expressed awareness and the extensive literature that exists on that, as outlined by Richardson (2021), managers frequently fail to make use of fairness frameworks and guidelines. Institutions can focus on making technical tools for fair AI more accessible and understandable for developers as well as help managers to effectively incorporate fairness frameworks into their organizations.

Additionally, the findings demonstrated that despite the prioritization of transparency, information about algorithms is not officially accessible to users through the platform. Policymakers could act on that by requiring platforms to have transparency sections on their websites, which would enable users to easily comprehend how algorithms make decisions.

6. LIMITATIONS AND FUTURE RESEARCH

The conducted research contributes to algorithmic fairness literature by providing a qualitative in-depth analysis presenting the decisions managers make surrounding the fair use of algorithms on online labor platforms. Nonetheless, this paper implies 3 limitations.

One of its main limitations is the fact that there were three participants, resulting in a small sample size in relation to standards in qualitative research. Due to unforeseen circumstances and the time constraints faced, the interviewing of additional employees was not possible. The small sample hinders the achievement of theoretical saturation, as it becomes difficult to fully investigate the research questions and arrive at a point where additional data collection is unlikely to produce new insights. Consequently, the generalizability of the results to a larger population may be constrained just as its ability to represent a wide range of perspectives and experiences. Moreover, the small sample also raises the possibility of sampling bias. The external validity and reliability of the results would be improved in future research with a larger and more varied sample.

The second limitation of this study is the exclusive focus on participants from Dutch online labor platforms. This geographic restriction constrains the generalizability of the findings to a wider international context. Future studies with participants from various geographic locations would contribute to a more in-depth understanding of the topic. Despite the geographical restriction, the study provides valuable insights into the Dutch online labor platform sector due to this area of research being novel and understudied. Additionally, the results of this study have some generalizability to other nations that are comparable to the Netherlands. According to the country similarity index, Germany, Belgium, and Denmark share a lot in common with the Netherlands, suggesting that the conclusions of this study may also be applicable to these nations (Jones, 2023).

A further limitation is the inclusion of both managers and software developers in this study. This introduces the possibility of biased and contradictory results due to their varying roles within the organization. According to the findings, developers frequently adopt a more technical viewpoint, whereas managers place more emphasis on the platform's overall fairness. Future research projects might address this issue by concentrating on people in comparable roles within the organization.

The unexpected finding that platforms outsource their algorithms limits the study's ability to fully explore the technical facets of algorithm design. Future research should focus on platforms that develop their algorithms internally in order to better understand the practices managers use to encourage fairness in algorithmic design. This would allow for a more thorough and nuanced analysis of the managerial tactics used to ensure the fair development of algorithms.

In addition to that, future research could delve into the underlying causes of managers' and developers' observed limited use of software toolkits, frameworks, and other solutions, and the reasons behind their preference for internal decision-making processes as observed in this study.

7. CONCLUSION

Despite the increased use of algorithms on online labor platforms, the existing literature failed to evaluate to what extent online labor platforms account for the potential bias-amplifying as well as controlling effects of algorithms. Drawing on theories from the fairness and the online labor platform realm, this research provided a small-scale exploratory study by examining managerial practices surrounding the fair design and deployment of machine learning algorithms on online labor platforms. By conducting semi-structured interviews with individuals holding managerial positions this study aimed at answering the following research question:

What do managers do to ensure the fair design and deployment of machine learning algorithms on online labor platforms?

This study expands upon the body of prior research in this area by providing insightful findings on how decisions are made, the motivations behind them, and the methods used to ensure fairness in algorithmic systems. The empirical findings and discussion showed that managers are aware of algorithmic as well as societal biases on their platforms and take various proactive approaches to mitigate the occurring biases. Moreover, it was found that despite the limited use of fairness tools as well as the lack of technical bias mitigation approaches in algorithm development, managers made a significant effort in ensuring platform fairness and showed general accountability for ensuring the fair design of algorithms. This exploratory study yielded additional results, which show that managers on online labor platforms frequently rely on internal team discussions rather than external frameworks when making decisions about the fair design and use of algorithms. Furthermore, the results illustrated the key challenges faced by managers in relation to algorithmic fairness. Overall, the findings highlight the complex nature of managing algorithmic fairness and underscore the need for further research and practical considerations in this area.

8. REFERENCES

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2477899

Berg, J. (2018). Digital labour platforms and the future of work: Towards decent work in the online world. *International Labour Organization*. https://www.ilo.org/global/publications/books/WCMS_645337/lang--en/index.htm

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Chakraborty, J., Majumder, S., & Menzies, T. (2021a). Bias in machine learning software: why? how? what to do? In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.1145/3468264.3468537

Chaudhuri, A., & Gangadharan, L. (2007). An Experimental Analysis of Trust and Trustworthiness. *Southern Economic Journal*, *73*(4), 959–985. https://doi.org/10.1002/j.2325-8012.2007.tb00813.x

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 6889–6892. https://doi.org/10.1073/pnas.1018033108

Daskalova, V. I. (2018). Regulating the New Self-Employed in the Uber Economy: What Role for EU Competition Law? *German Law Journal*, *19*(3), 461–508. https://doi.org/10.1017/s207183220002277x

Dastin, J. (2022). Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women *. In *Auerbach Publications eBooks* (pp. 296–299). Auerbach Publications. https://doi.org/10.1201/9781003278290-44

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, *30*(1), 114–132. https://doi.org/10.1111/1748-8583.12258

Edelman, B., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2377353

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, *62*(4), 379–384. https://doi.org/10.1007/s12599-020-00650-3

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Galperin, H., & Greppi, C. (2017). Geographical Discrimination in Digital Labor Platforms. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2922874

Garcia-Gathright, J. (2018, September 10). *Assessing and Addressing Algorithmic Bias - But Before We Get There*. arXiv.org. https://arxiv.org/abs/1809.03332

Hannak, A., Wagner, C., Garcia, D. A., Mislove, A., Strohmaier, M., & Wilson, C. (2017). Bias in Online Freelance Marketplaces. In *Conference on Computer Supported Cooperative Work*. https://doi.org/10.1145/2998181.2998327

Howard, A. M., & Borenstein, J. (2018). The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics*, *24*(5), 1521–1536. https://doi.org/10.1007/s11948-017-9975-2

Jahanbakhsh, F., Cranshaw, J., Counts, S. E., Lasecki, W. S., & Inkpen, K. (2020). An Experimental Study of Bias in Platform Worker Ratings: The Role of Performance Quality and Gender. In *Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376860

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Jobin, A., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jones, J. (2023). The Most Similar Countries to the Netherlands. *OBJECTIVE LISTS*. https://objectivelists.com/2022/06/14/which-countries-are-most-similar-to-netherlands/#:~:text=According%20to%20the%20Index%2C%20Belgium,native%20speakers%20of%20Germanic%20languages.

Jordan, M. I., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kim, J., & Cho, S. (2022). An information theoretic approach to reducing algorithmic bias for machine learning. *Neurocomputing*, *500*, 26–38. https://doi.org/10.1016/j.neucom.2021.09.081

Lum, K., & Isaac, W. S. (2016). To predict and serve? *Significance*, *13*(5), 14–19.

Mattu, J. a. L. K. (2020, February 29). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing https://doi.org/10.1111/j.1740-9713.2016.00960.x

Mehrabi, N. (2019, August 23). *A Survey on Bias and Fairness in Machine Learning*. arXiv.org. https://arxiv.org/abs/1908.09635

Meijerink, J. G., Keegan, A., & Bondarouk, T. (2021). Having their cake and eating it too? Online labor platforms and human resource management as a case of institutional complexity. *International Journal of Human Resource Management*, *32*(19), 4016–4052. https://doi.org/10.1080/09585192.2020.1867616

Mohlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control. *Management Information Systems Quarterly*, *45*(4), 1999–2022. https://doi.org/10.25300/misq/2021/15333

Monachou, F. G. (2019). *Discrimination in Online Markets: Effects of Social Bias on Learning from Reviews and Policy Design*. https://proceedings.neurips.cc/paper/2019/hash/e00406144c1e7e35240afed70f34166a-Abstract.html

Mukerjee, A., Biswas, R., Deb, K., & Mathur, A. (2002). Multi-objective Evolutionary Algorithms for the Risk-return Trade-off in Bank Loan Management. *International Transactions in Operational Research*, *9*(5), 583–597. https://doi.org/10.1111/1475-3995.00375

Olteanu, A., Castillo, C. F., Diaz, F., & Kiciman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*. https://doi.org/10.3389/fdata.2019.00013

Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Tal, A. S., Hartmann, A.,

& Kuflik, T. (2022). Mitigating Bias in Algorithmic Systems—A Fish-eye View. *ACM Computing Surveys*, *55*(5), 1–37. https://doi.org/10.1145/3527152

Park, S., & Ryoo, S. (2023). How Does Algorithm Control Affect Platform Workers' Responses? Algorithm as a Digital Taylorism. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(1), 273–288. https://doi.org/10.3390/jtaer18010015

Proudfoot, K. (2022). Inductive/Deductive Hybrid Thematic Analysis in Mixed Methods Research. *Journal of Mixed Methods Research*, *17*(3), 308–326. https://doi.org/10.1177/15586898221126816

Richardson, B. (2021, December 10). *A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions*. arXiv.org. https://arxiv.org/abs/2112.05700

Rieke, M. B. A. (2018, December 9). *Help wanted: an examination of hiring algorithms, equity, and bias*. https://apo.org.au/node/210071

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing Bias in AI. In *The Web Conference*. https://doi.org/10.1145/3308560.3317590

Saxena, N. (2019). Perceptions of Fairness. In *National Conference on Artificial Intelligence*. https://doi.org/10.1145/3306618.3314314

Sekiguchi, K., & Hori, K. (2020). Organic and dynamic tool for use with knowledge base of AI ethics for promoting engineers' practice of ethical AI design. *AI & Society*, *35*(1), 51–71. https://doi.org/10.1007/s00146-018-0867-z

Spitko, E. G. (n.d.). *Reputation Systems Bias in the Platform Workplace*. BYU Law Digital Commons. https://digitalcommons.law.byu.edu/lawreview/vol2019/iss5/7/

Stanford, J. (2017). The resurgence of gig work: Historical and theoretical perspectives. *Economic and Labour Relations Review*, *28*(3), 382–401. https://doi.org/10.1177/1035304617724303

Vanover, C., Mihas, P., & Saldana, J. (2021). *Analyzing and Interpreting Qualitative Research: After the Interview*. SAGE Publications.

Webster, C. S., Taylor, S., Thomas, C. S., & Weller, J. (2022). Social bias, discrimination and inequity in healthcare: mechanisms, implications and recommendations. *BJA Education*, *22*(4), 131–137. https://doi.org/10.1016/j.bjae.2021.11.011

Wohlin, C. (2014). *Guidelines for snowballing in systematic literature studies and a replication in software engineering*. https://doi.org/10.1145/2601248.2601268

APPENDIX
Appendix A – Interview Protocol

Introductory questions

1. Who are the users of your platform?
2. What industry does your organization operate in?
3. What is the business model? What is the value proposition/service your platform offers these users? In what way does your organization generate revenue?
4. What is the size of your organization in terms of employees and revenue?
5. What is your educational background, and what is your prior job experience?
6. How long have you been working at the organization?
7. What are your main tasks and responsibilities?
8. In your role as a [software developer, designer, or manager of a team of designers], what decisions can you make independently, and what decisions are made for you?
9. Which other individuals within and outside the organization do you work together with or depend on?
10. For which main decisions or features of your platform are learning and/or automating algorithms used?

Algorithms & Fairness

1. Have you ever encountered any type of discrimination/unfair treatment of workers on your platform? Can you name an example?
2. What role do you think the use of algorithms plays in reinforcing the stated bias/discrimination?
3. What does fairness mean for you?
    a) How would you define fairness in an organizational context?
    b) Do you have an ethics/fairness framework? If not, why not? If yes, what does it say? /(Do you have fairness boards/committees/organizational fairness structure?)
4. How do you ensure that your/(and/or) company's fairness values are embedded into the development of an algorithm (in each stage of the process)? How do you reduce the stated biases that might occur (in each stage)?
    a) What do you do in the data collection stage? How? Why?
    b) What do you do in the modeling/algorithm design stage? How? Why?
    c) What do you do in the deployment stage? How? Why?
5. How do you decide whether an algorithm is fair/safe to use? Who decides that? Do you use any framework/checklists for that?

6. How do you make sure that developers reflect your company's fairness values in the system?

7. What guidelines/metrics/systems/procedures/frameworks do you provide to developers to ensure that the systems they develop are fair? Do you use software toolkits or checklists? If yes, which ones?

8. What challenges do you face when ensuring the fair design/use of algorithms?

9. How do you make sure developers prioritize fairness over organizational goals or their own careers?

Appendix B – Interviewee background information

Participant	Gender	Platform	Experience & position at the company	Prior Education
E1	Male	8Vance.com	June 2018 – November 2020: -AI and Data Science Specialist November 2020-Present: -R&D Lead	Msc Artificial Intelligence at Maastricht University
E2	Male	8Vance.com	February 2022 – Present: - AI Engineer	Msc Data Science at Eindhoven University
E3	Male	Babysits.nl	March 2008 – Present: -Founder & CEO	Msc, Entrepreneurship and New Business Venturing at Erasmus University Rotterdam School of Management

Appendix C – Platform interface

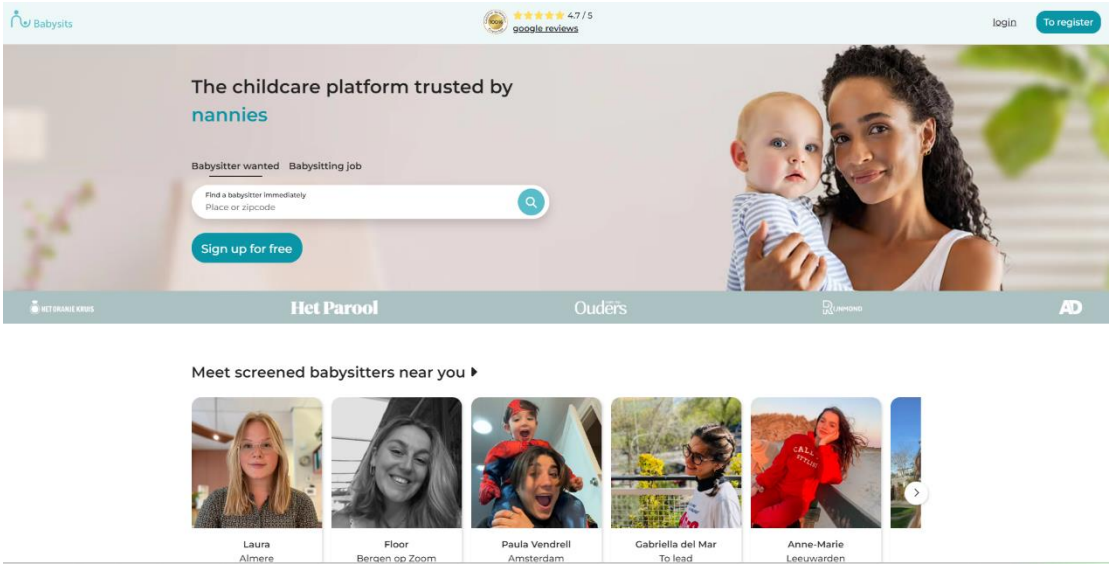

Figure 2.1: Screenshot of Babysits Website interface; extracted on 28.06.2023 (https://en.babysits.nl/)

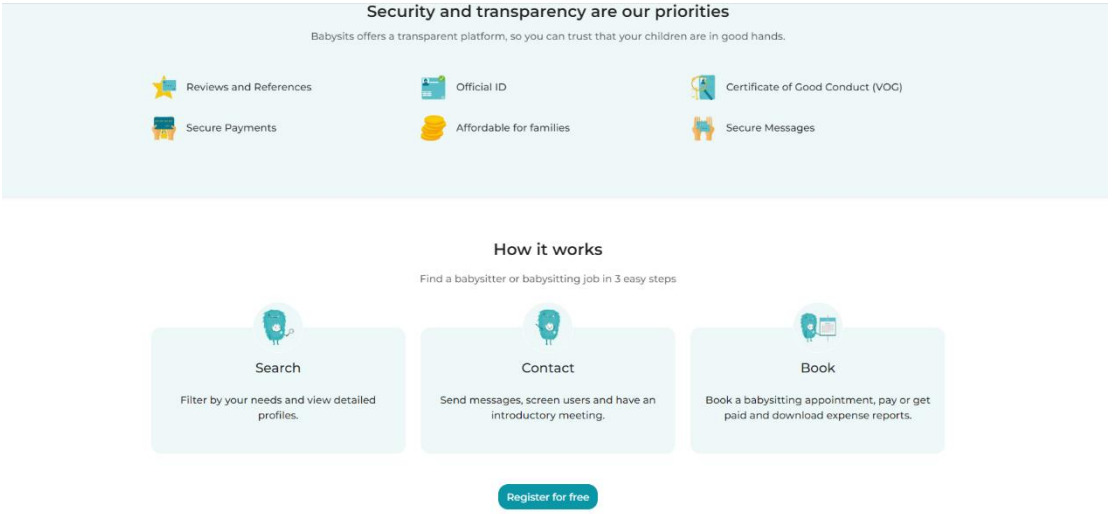

Figure 2.2: Screenshot of Babysits Website interface; extracted on 28.06.2023 (https://en.babysits.nl/)

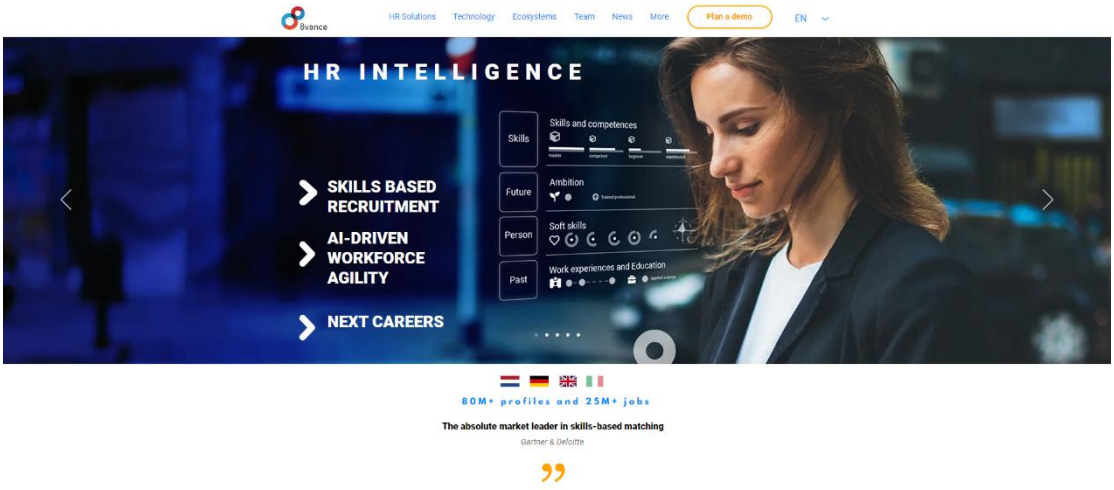

Figure 3.1: Screenshot of 8Vance Website interface; extracted on 28.06.2023 (https://www.8vance.com/?lang=de)

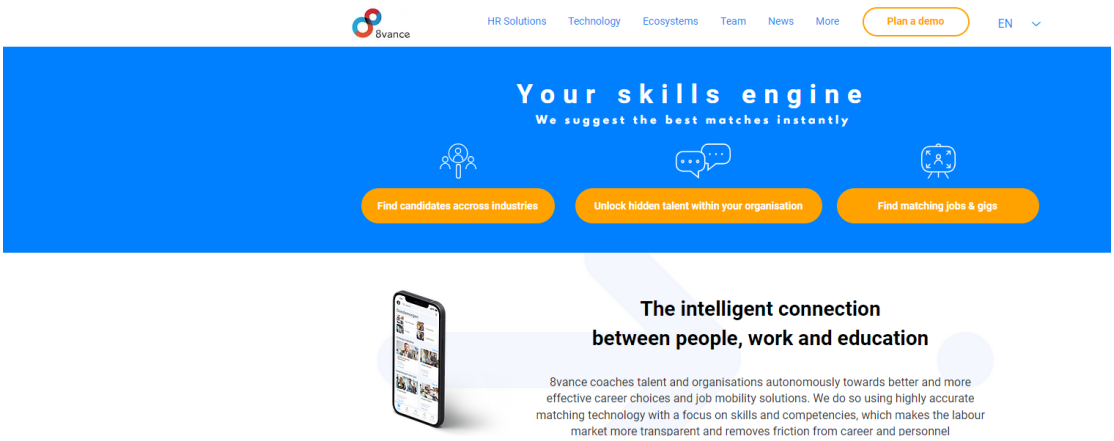

Figure 3.2: Screenshot of 8Vance Website interface; extracted on 28.06.2023 (https://www.8vance.com/?lang=de)