A Work Project, presented as part of the requirements for the Award of a Master Degree in Management from the NOVA – School of Business and Economics.

ARTIFICIAL INTELLIGENCE IN RECRUITMENT: JUST BECAUSE IT'S BIASED, DOES IT MEAN IT'S BAD?

GIUSY BENEDUCE | 39652

Work project carried out under the supervision of: Professor João Castro

22nd May 2020

Abstract

In recent years, artificial intelligence technologies have seen a major diffusion. Despite the numerous benefits, some concerns arise around the possibility that artificial intelligence may inherit human bias. This paper addresses to what extent artificial intelligence bias is a blocker to the implementation of those technologies in recruitment. The research was developed focusing on gender bias and data was gathered through a survey-based questionnaire. Data shows that recruiters consider artificial intelligence an opportunity to improve the recruitment process, however, bias does represent a disincentive to the implementation of those technologies.

Keywords: Artificial Intelligence, Recruitment, Bias, Gender discrimination

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

1. Introduction

In recent years, artificial intelligence-based technologies have seen a major development. The diffusion of these technologies in sensitive sectors such as healthcare, criminal justice and human resources have generated a keen interest from experts and general public. (Silberg and Manyika 2019). In particular, many researchers are studying the ethical impact that the use of artificial intelligence in certain sectors can have on society.

In this regard, many questions arise: what does it mean that artificial intelligence systems make decisions? What are the ethical and moral consequences on society? Who is accountable for artificial intelligence systems' decisions? Will decisions taken by artificial intelligence systems be less biased than human ones? Or will they only exacerbate existing biases?

It is clear that a new, ethical challenge is emerging for artificial intelligence. When algorithms engage in cognitive performances, which are by definition carried out by humans, at least until now, they inherit the compliance to social requirements. This outlines a challenge for companies that must ensure transparency, fairness and respect for the founding values of our society (Bostrom and Yudkowsky 2014).

This work project will try to address the artificial intelligence bias issue with a focus on gender bias in the recruitment sector. There are three major contributions. The first one is an overview, by no means exhaustive, of the literature written so far about general algorithmic fairness and artificial intelligence bias. The second one is a picture of the impact that artificial intelligence is having on the recruitment industry and the relative challenges. The third one revolves around an empirical study that aims to answer the following question: to what extent is bias a blocker to the implementation of artificial intelligence solutions in the recruitment sector?

Due to the lack of academic literature about the influence of artificial intelligence on the recruitment industry, and its impact on gender bias, both scholarly literature and professional sources were used.

2. Literature review

2.1. Artificial intelligence and fairness

To fully understand the subject, it is necessary to define what is meant by "artificial intelligence" and what is meant by "fairness" and “bias”. The term artificial intelligence (AI) was coined by John McCarthy (2007) and it was defined as “the science and engineering of making intelligent machines”. It’s a broad definition that includes many different types of AI. One of the most known ones is machine learning. It can be defined as “an automated process of discovering correlations between variables in a dataset, often to make prediction or estimate some outcome” (Lerh and Ohm 2017). In this study, the term “AI” refers to any type of artificial intelligence application.

Fairness is generally defined by the Oxford Dictionary as an impartial behavior, however, defining algorithmic fairness is much more challenging. The computer scientist Arvind Narayanan (2018) listed 21 fairness definitions and stated “any overarching definitions will inevitably be vacuous” because to achieve one measure of fairness means to give up on others and therefore is not possible to satisfy them all. In another research conducted by Kleinberg, Mullanathan and Raghavan (2016), a model to score the risk class for credit decision was analyzed and three definitions of fairness were developed. The authors demonstrated that it was impossible to satisfy all three definitions simultaneously and therefore the existence of a trade-off in fairness definition. Against this background, it is possible to say that fairness isn’t a unique concept, but it must be considered according to the context and the stakeholders to which it applies. The AI Now Institute (2019) also tackled the limitations of a narrow fairness definition and to respond to this problem, many researchers refer to the "causal fairness methods". Those approaches aim to identify causal relationships between different types of data and the various outcomes they produce. Therefore, AI is considered fair when factors such as race or gender do not causally influence the model's prediction (Crawford 2019). In this study, the words "fair" or "unfair" are to be considered according to the concept of "causal fairness". Finally, for the sake of clarity, the term "bias" is to be considered within the meaning of "undesirable bias" that is described as systematic discrimination against certain individuals or group of individuals based on the inappropriate use of certain traits or characteristics (Friedman and Nissembaum 1996).

2.2. Algorithmic bias and discrimination risk

AI has the potential to be the antidote to human bias and to improve fairness in many aspects of our lives. But it would be incorrect to think that algorithms reveal an objective truth just because they are mathematical. The AI bias problem lays in the algorithms' incapability to understand the social and historical context of the data they use (Crawford 2019). This is extremely difficult to achieve and what emerges is the so-called "bias in, bias out" phenomenon. This means that inequality and discrimination lie in the nature of prediction. Making a prediction about future events looking at the past, will project the inequalities of the past into the future (Mayson 2018). Recently, many studies have been aimed at showing that AI systems inherit human and societal bias with the consequence of exacerbating discrimination against minorities. In 2016, ProPublica conducted an analysis on COMPAS, an AI tool used, among others, in the USA to predict a criminal defendant's likelihood of becoming a recidivist. The main finding was that black defendants were more likely to be incorrectly judged than white defendants. Data shows that 45% of black defendants classified as high risk did not recidivate over a two-year period against 23%

for white defendants. Black defendants who re-offended within two years and that were wrongly judged low risk were 28% against 48% for white defendants. Even though COMPAS did not directly include the variables race, the tool discriminated against the black defendants (Larson, 2016).

In a work published in 2018, Joy Buolamwini demonstrates that automated facial analysis algorithms discriminate based on categories such as race and gender. The study was conducted on gender classifiers sold by IBM, Microsoft and Face ++. The research shows that all classifiers perform better on male faces than female faces with an 8,1% error rate for males and 20,6% for females. They also perform better on lighter complexion faces with an error rate of 11,8%, while the error for darker complexion faces is 19,2%. The result was that darker-skinned females were the most inaccurate group demonstrating a substantial disparity in the accuracy of classifying males and females and darker and lighter-skinned people (Buolamwini and Gebru 2018).

Those researches, among the others, demonstrate that AI can absorb human bias and perpetrate the same inequities that it was supposed to solve and therefore show how AI used in sensitive areas is potentially harmful.

2.2.1. Discrimination causes

The reason of AI bias may lay in the training data, but it would be simplistic to think that it is the only cause of AI bias. AI decision-making can lead to discrimination in many other ways. Barocas and Selbsts (2016) examined several causes of AI discrimination, namely the definition of the target variables and class labels, labeling and collection of training data, features selection, proxies and intentional discrimination.

In AI, data mining is the extraction of implicit, previously unknown and potentially useful information from data. It automates the process of discovering useful patterns and the set of relationships found is a defined model, which is used to make predictions or estimations of unobserved variables (Hand 2012). The definition of the target variable is not an obvious task. The data scientist has the role to translate a conceptual problem in formal terms so that it can be analyzed by a computer. This is a subjective process that can lead to a definition of the problem in such a way that unintentionally disadvantage protected classes (Custers 2013). Training data can be defined as the data that trains the model to behave in a certain way. It can also be a source of discrimination, in fact when prejudice happened in the past, training data results to be biased and the model learns from a pre-existing prejudice and simply reproduce it (Berry 2004). Another case is when the sample population is biased. In this case, a group of people may be under-represented or over-represented in the dataset and the model could systematically create a disadvantage for them (Barocas and Selbsts 2016). The feature selection is the process of choosing what attributes to observe conducting the analysis. (Liu and Motoda 1998). These decisions might also have an impact on protected classes and create discrimination. For example, if an organization wants to predict which job applicants will be good employees, it normally focuses on certain features of the candidate. By doing so, bias against some groups of people may be introduced. As a matter of fact, recruiters often look for people who studied at famous universities, but it might be relatively rare for certain racial groups to study at those expensive universities. Therefore, it may have discriminatory effects if an employer selects job applicants based on the university they had chosen to study (Barocas and Selbsts 2016). Another source of bias are “proxies”. They are defined as "criteria that are genuinely relevant in making rational and well-informed decisions and also happen to serve as reliable proxies for class membership" (Barocas and Selbsts 2016). This problem is called "redundant encodings” and happens when the belonging in a protected class is encoded in other data (Dwork et al. 2012). Therefore, some criteria that are important for the outcome of the analysis could also be a source of discrimination and thus create situations in which data scientists, even with the aim of conducting an accurate analysis, end up discriminating against certain protected categories (Barocas and Selbsts 2016). Finally, discrimination could also occur intentionally, and data mining could also provide a new way to disguise the intention to discriminate. In fact, decision-makers could deliberately influence data collection to ensure that mining suggests less favorable rules for members of the protected classes (Barocas and Selbsts 2016).

2.2.2. How to achieve algorithmic fairness

Avoiding bias completely is nearly impossible. It is something intrinsic to the very concept of fairness, which has many shades and therefore there is no univocal solution to the problem (Mehrabi 2019). Currently, there are no regulations on the subject, but like the issue of privacy and data treatment, ethics in AI is a relevant topic that will need specific regulations (Tambiama, 2019). To date, however, many public and private organizations have drawn up the guidelines for ethical use of AI. In a research published in 2019 Jobin, Ienca and Vaiena analyzed 84 documents containing ethical principles or guidelines for AI. Although there is no global agreement, the guidelines developed by different organizations converge in five key requirements that AI systems need to meet in order to be deemed trustworthy: transparency, justice and equity, non-maleficence, responsibility and accountability, privacy. In order to ensure transparency, companies should disclose information about those developing or deploying AI systems and the information should be explained in a manner adapted to the stakeholder concerned. Justice and equity are mainly expressed in terms of prevention, monitoring and mitigation of unwanted bias, discrimination and marginalization of vulnerable groups. Non-maleficence is defined as security that AI should never cause foreseeable harm. Then, in order to be responsible and accountable companies should enable the assessment of algorithms, data and design processes. Finally, companies should ensure full respect for privacy and data protection, use adequate data governance mechanisms and ensure a legitimized access to data.

2.3. Artificial intelligence in recruitment: State of the art

Recruitment is a sector that lends itself well to be revolutionized by AI, given its structural characteristics. In fact, activities such as searching for candidates, screening CVs or matching job descriptions can be automated and made more efficient by systems that can analyze a massive amount of data, run correlations and make data-driven decisions (Ahmed 2018). Because of the enormous benefit that AI promises to recruitment, and because it is a sensitive sector, a marked interest has arisen around its adoption and its consequences on the job market. In this section, the author wants to illustrate the state of the art of AI in recruitment as well as to define the main advantages and concerns. Finally, a focus on the problem of gender bias in recruitment is presented including three cases of AI gender discrimination.

2.3.1 The impact of artificial intelligence in recruitment

The hiring process is a series of activities that can be summarized in four main stages: sourcing, screening, interviewing, selection (Bogen and Rieke 2018). Each of these stages includes several different activities, and AI tools can change the way each of these phases is carried out.

The first stage in the recruiting process is sourcing candidates. Sourcing refers to the phase where recruiters seek out candidates with the desired characteristics to apply for their job opportunities. It includes writing a job description, advertising the job opening and headhunting (Bogen and Rieke 2018). Starting from the job description, there are currently tools helping recruiters to write a more effective job description depending on the type of profiles they want to attract (Bogen and Rieke 2018). The technology behind them relies on comparing the linguistic pattern in the text with the applicant response and the hiring outcome to predict the characteristics of the pool of candidates that is likely to respond to the job posting (Textio company website n.d.). Advertising is a crucial activity in the selection process because it determines the pool of candidates. Just like any other online business that directs its advertisements to a specific target audience, there are several platforms that help recruiters to target the desired profiles. The most striking example is social media platforms that allow targeting their users based on data such as demographics, search terms, geographic location, interests and social contents (Dalenberg 2018). While the employer chooses the target audience that she wants to reach, social media platforms make their own predictions, on which users are more likely to engage with the ad and, based on that, decide which users to show the ad (Bogen and Rieke 2018). Headhunting is the action of proactively identifying and approaching desirable candidates and it is common for companies to look for candidates with specific skills or qualifications. (Bogen and Rieke 2018). There are tools on the market enabling recruiters to search for different sources and collect the most suitable profiles, as well as make predictions based on them. It is in fact possible to estimate a company's health, the probability of a candidate leaving the current job or having certain skills even if not explicitly indicated (Entelo company website n.d.).

Once recruiters have the pool of candidates, the second stage is the screening, which consists of reviewing job applications, and involves reading resumes and cover letters to find the best profile. This phase includes the evaluation and the appraisal of candidates' qualifications and AI tools can help recruiters to score and rank applicants and assess their capabilities (Bogen and Rieke 2018). Many companies offer tracking systems in order to screen applications based on a list of predefined keywords and to shortlist the resumes that better match with the job description and appear to meet the requirements. Some systems can also rank the candidates and present them by order of relevance based on qualification indicators (Mya company website n.d.). Other screening tools aim to evaluate whether candidates are fit for the job. This can be achieved through tests and games that are designed to assess their cognitive and logical capabilities, social and emotional behavior, memory, processing speed in order to predict their future job performance (Ahmed 2018).

The next stage is the interview. This is probably the most personal phase of the selection process and therefore less likely to be automated using AI. Nevertheless, there are tools that allow recruiters to carry out video interviews and analyze candidates' responses, facial expressions and the tone of voice (Ahmed 2018). The last phase is the selection, where employers make final hiring decisions. It usually includes background checks and negotiations of terms. In this phase, AI tools help to predict candidates' likelihood of violating workplace policies as well as estimating remuneration and company benefits (Bogen and Rieke 2018).

2.3.2. Implementation of artificial intelligence in recruitment: Drivers

According to pieces of research, there are three key drivers for the implementation of AI in recruitment: hiring time, hiring quality and the cost of the hiring process. (Bogen and Rieke 2018) (Linkedin Talent Solutions 2018). Time is a crucial factor in the hiring process and "time to hire" is the most common recruiting measure because it tracks the immediate outcome of recruiters' efficiency (Linkedin Talent Solutions 2019). In fact, the recruiting process involves time-consuming activities such as sourcing candidates, screening applications or conducting interviews that can make the selection process long and costly. *Ideal*, an AI company operating in the recruitment sector, estimated that screening resumes and shortlisting candidates for interviews take 23 hours of a recruiter's time for a single hire, while the average period of the time needed to hire a single candidate is 42 days (SHRM 2016). It follows that the use of AI systems has a huge impact on the time needed to hire a candidate.

The quality of hire is one of the most common measures used by companies because it refers to the long-term business impact of new employees. Most businesses define the quality of hiring as a combination of three core factors: retention, engagement, and performance ratings (Linkedin Talent Solutions 2019). AI tools may help to improve the quality of the candidate hiring process in several ways, starting from an accurate job description and targeted advertisement, all of which attract adequate applicants (Okolie 2017). Moreover, automated CVs screening systems allow recruiters to consider a larger pool of candidates that otherwise would be ignored. In fact, when the screening is handled manually, recruiters are forced to limit their review of the applicant pool, that can mean ignoring potential talents (Ideal company website n.d.).

Both saving time and the improved quality of hire drive are cost-effective. The cost of a new hire includes hours spent reviewing resumes and interviewing candidates and recruitment advertising fees. It also includes the time and expenses associated with onboarding & training new employees (Hire by Google 2019). Poor performance and low retention rate mean the loss of the investment for the company. Indeed, poor hiring decisions are estimated to cost $1.6 million for every 1,000 hires made (Wright and Atkinson 2019). Consequently, despite the initial investment that the implementation of AI systems may entail, these can lead to substantial cost savings.

2.3.3. Implementation of artificial intelligence in recruitment: Blockers

According to a study conducted by HR.com (2019), there are two most common concerns about the AI implementation in recruitment: the lack of human factor and the exacerbation of human bias. The lack of human factor from the point of view of the candidate may translate into a poor hiring experience, while from the recruiters' point of view it could mean giving up on some aspects of the assessment (Sanderson n.d.). The second most common concern is the perpetuation and even the exacerbation of human bias. AI learns from humans and can inherit human and social prejudices with the result of perpetuating discrimination against minorities and this could represent a blocking factor to a wide implementation of those technologies (Lloyd 2018).

2.4. Artificial intelligence exacerbation of gender bias in recruitment

Despite women's advancement in the workplace, the gender gap remains a current issue. Globally, only 36% of senior private sector's managers and public sector's officials are women and the gap becomes even wider when it comes to science, technology, engineering, and math (STEM) careers (World Economic Forum 2020). Several studies show that there is a tangible risk of gender stereotypes influencing recruitment through AI. Below are three cases in which it has been demonstrated that AI applied to the field of job search has produced discrimination against women.

2.4.1. The Google case

In 2015 three researchers from Carnegie Mellon University and the International Computer Science Institute conducted an experiment on the Google advertisement system. The goal of the experiment was to determine in what circumstances belonging to certain categories causes significant changes to ads by comparing many instances of each social group. One of the categories analyzed was gender. The experiment was conducted using a tool that collects ads served by

Google and automatically analyzes the data to determine whether statistically significant differences between groups of people exist. Both male and female visited webpages related to the topic of employment and the result was that Google showed male ads from a certain career coaching agency promising high salaries more frequently than the female ads did. There is no evidence that the discrimination was intentional, and no laws or policies were broken. Instead, it is believed that the discrimination might have resulted from click-through rates optimization performed by algorithms (Datta, Admit, Datta 2015).

2.4.2. The Facebook case

In 2018 two researchers from the London Business School and the MIT Sloan School of Management conducted a research on how Facebook algorithm delivers ads promoting job opportunities in STEM fields. The test was conducted on Facebook in partnership with a website that gives information about STEM careers. Advertising campaigns would direct users who clicked on the ad on Facebook to that website. The ad was targeted at both male and female adults in 191 countries. The results show that women were less likely to be shown the ad, compared to men. But not because they were less likely to click on it. In fact, when women saw the ad, they were more likely than men to click. The gender-imbalance is instead due to the fact that women are a prized demographic and as a consequence are more expensive to show ads to. This leads to the conclusion that an ad algorithm which optimizes ad delivery to be cost-effective, can deliver ads, intended to be gender-neutral, in a discriminatory way (Lambrecht and Tucker 2018).

2.4.3. The Amazon case

In 2014, Amazon set up a tool to review job applicant's resumes that used the natural language process and machine learning to hire the top applicants. The idea was that the software would use AI algorithms to learn key traits from successful job applicants' resumes over a period of time and look for similar markers in resumes submitted for screening. The tool would then rate the candidate on a scale from one to five, depending on how closely they resemble prior successful candidates. Almost one year after the use of this tool, the company found out that for technical jobs such as software developers and architects, the ratings were not done in a gender-neutral way. Amazon came to the conclusion that the cause of the bias was the data used for training the AI system, consisting mostly of male employees' resumes. Therefore, the biased training data led the algorithms to downgrade resumes that included words like "women's". This discovery forced Amazon to withdraw the tool and design new algorithms to be unbiased (Kodiyan 2019).

3. Motivation and research question

The literature presented above shows that, on the one hand, AI has a great potential to change the way selection processes are carried out today. On the other hand, it was proved that AI can inherit human bias and lead to discrimination in sensitive sectors such as recruitment. Although the risk of algorithmic bias has been demonstrated, it has never been investigated whether this is actually a problem for companies and whether this is an impediment to the implementation of AI systems. From an ethical point of view, bias is directly perceived as negative. But do companies think ethically?

In this regard, this research aims to answer the following question: to what extent is bias a blocker to the implementation of artificial intelligence solutions in the recruitment sector?

4. Methodology

To collect data, a quantitative method was chosen in order to gather more objective and numerous answers. A survey was formulated and distributed to recruiters and human resources professionals. All respondents were personally contacted on Linkedin and invited to participate to the research, then the survey was personally sent to them.

The survey consisted of three sections. The first section aimed to investigate the respondent's general point of view on bias in recruitment. To do so, six statements regarding unconscious bias in recruitment, and the possibility to remove it, were presented. Respondents were asked to reveal to what extent they agreed to each statement.

The second section aimed to investigate what recruiters think about the use of AI in recruitment. To do so, five questions were presented. Respondents were asked to pick an alternative between two statements about the risks and benefits of AI in recruitment.

The third section aimed to discover to what extent bias is a deterrent to the implementation of AI in recruitment. To do so, twelve questions were presented. The three main advantages of AI in recruitment were considered: hiring quality, hiring time and the cost of the hiring process. Respondents were asked to declare to what extent they were willing to sacrifice these three factors in order to have an unbiased selection process. Moreover, for each factor, two hypothetical scenarios have been designed.

5. Presentation of results

5.1. Sample

In total 158 responses were collected. Firstly, a missing values analysis was performed. The main reason for missing values was that respondents didn't finish the survey. Since the number of missing values was higher than 10% and given the non-randomness of the missing values, incomplete observations were excluded in order to preserve the representativeness of the database. Therefore, the analysis was performed on a sample of 101 observations. 52.5% of respondents were female and 47.5% were male. The majority of respondents (60.4%) was aged between 25 and 35 years, followed by people younger than 25 years (19.8%), people aged between 35 and 45 years (13.9%) and people aged between 45 and 55 years (5.9%). The first country of origin of respondents was Italy (40.6% respondents), followed by Portugal (18.8%), Spain (10.9%) and Germany (10.9%). All other countries accounted for less than 4% of the total respondents. Most of respondents were working in consulting (28.7%), information technology (11.9%), healthcare

(7.9%), consumer goods (7.9%), banking (5.9%). All other industries accounted for less than 5% of respondents.

5.2. Attitude towards bias in recruitment

The survey results show that, overall, recruiters and human resources experts are aware of the issue of bias in recruitment. In fact, 23,8% of them agreed and 57,4% of them partially agreed to the statement *"Hiring processes are always subjected to unconscious bias".* Data shows that 18,8% of respondents agreed to the statement *"Bias in hiring processes is hard to remove"*, while 51,5% partially agreed. Also, 38.6% of respondents agreed to the statement *"I'd like my company to put more effort to reduce bias in hiring processes".* However, many respondents are not aware of the measures taken by their companies to reduce bias in recruitment. 36.6% declared their selves neutral to the statement *"My company implements measures to reduce bias in hiring processes"*. Moreover, bias is not perceived as something absolutely negative, in fact 32.7% of respondents stated to be neutral to the statement *"Bias does not necessarily affect hiring processes in a negative way"*, while 26,7% partially agreed and 15,8% agreed. Finally, technology is considered a potential solution to reduce bias, in fact 41,6% of respondents partially agreed and 22,8% agreed to the statement *"Bias in hiring processes can be reduced by using technology".*

5.3. Attitude towards artificial intelligence in recruitment

Overall, respondents have shown a positive attitude towards the use of AI in selection processes. In fact, 79,2% of respondents stated that "*Artificial intelligence improves the quality of recruitment"* and 86,1% believed that *"Artificial intelligence decreases bias in the recruitment process"*. Moreover, when asked to pick between the alternatives *"Human bias is easier to correct"* and *"Artificial intelligence bias is easier to correct",* 72,3% chose the second alternative.

However, despite the positive attitude, the results show that trust in AI still presents some reservations. While 39,6% of respondents believed that “*The lack of human judgement is an advantage of artificial intelligence tools”*, 60,4% stated that “*The lack of human judgement is a disadvantage of artificial intelligence tools”*. Also, 48,5% said that “*A human being is more likely to make a bad decision”*, while 51,5% believed that “*An artificial intelligence tool is more likely to make a bad decision”*, indicating some discordance on the subject.

5.4. Quality improvement versus bias

The quality of recruitment appeared to be a crucial metric for recruiters. For the sake of clarity, in the survey it was defined as *“the value a candidate can bring to the organization which can be measured with metrics like performance appraisal score and retention rate.”* Data shows that 57,4% of respondents said that they would not give up on the quality of the selection process in order to guarantee the absence of bias. The remaining 42,6% stated that they would sacrifice quality in order to eliminate bias in the selection process. Among those, 4 respondents would give up more than the 70% of recruitment quality, 11 respondents would give up between 50% and 70% of recruitment quality, 16 respondents would give up between 30% and 50% of recruitment quality, 11 respondents would give up between 10% and 30% and 4 respondents would give up less than 10% of recruitment quality. As for evaluating how recruiters deal with selecting between improving the quality and reducing the bias of the selection process, a hypothetical scenario was presented: *“Suppose that the company you are currently working for is thinking to acquire an artificial intelligence tool that would improve the quality of the selection process by forecasting the candidate's likelihood of having a high-performance appraisal score and the candidates' likelihood of staying in the company at least 5 years”*. Then, respondents were asked if they would acquire the tool knowing that it had a certain level of bias against women. On the assumption that the tool selected 70% male profiles and 30% female, 57,4% of respondents stated that they would prefer not to acquire the tool. On the assumption that the tool selected 60% male profiles and 40% female profiles, 39,6% of respondents stated that they would prefer not to acquire the tool. Therefore, from a total of 58 respondents who initially stated that they would sacrifice quality in order to reduce bias, 18 considered the second level of bias (60% male and 40% female) acceptable and opted for the acquisition of the tool.

5.5. Time efficiency versus bias

In order to evaluate the importance of the metric time, as for the quality, a trade-off was presented to recruiters. Data shows that 38,6% of respondents said that they would not give up on the time efficiency of the selection process in order to guarantee the absence of bias. The remaining 61,4% stated that they would sacrifice time efficiency in order to eliminate bias in the selection process. Among those, 2 respondents would accept a recruitment process up to 70% longer in order to eliminate bias, 11 respondents would accept a process between 50% and 70% longer, 22 respondents would accept a process between 30% and 50% longer, 22 respondents would accept a process between 10% and 30% longer and 5 respondents would accept a process less than 10% longer. Also for this metric, a hypothetical scenario was presented: "*Suppose you are the head of the HR department of a company and you have to acquire an artificial intelligence tool to improve the performance of your department.*" First, respondents were asked to choose between *"a tool that takes 1 week to select a new employee and statistically selects 70% men and 30% women, given the same number of male and female applicants"* and *"a tool that takes 3 weeks to select a new employee and statistically selects 60% men and 40% women, given the same number of male and female applicants"*. Results show that 71,3% of respondents preferred to have a three times longer process, instead of a more biased one. Then the same question was asked with a different level of bias and different process duration. Respondents had to choose between: *"A tool that takes 1 week to select a new employee and statistically selects 60% men and 40% women, given the same number of male and female applicants"* and *"A tool that takes 5 weeks to select a new employee and statistically selects 50% men and 50% women, given the same number of male and female applicants".* In this case, 48,5% of respondents chose to have a five times longer, but totally unbiased process. Therefore, out of 72 respondents that gave up on time efficiency in the first question, 23 considered the level of bias "60% men and 40% women" acceptable and opted for a more time-efficient process.

5.6. Cost reduction versus bias

Data shows that 85,1% of respondents said that they would give up on reducing the cost of recruitment in order to guarantee the absence of bias. Among those, 5 respondents would accept a recruitment process up to 70% more expensive to eliminate bias, 13 respondents would accept a process between 50% and 70% more expensive, 32 respondents would accept a process between 30% and 50% more expensive, 27 respondents would accept a process between 10% and 30% more expensive and 9 respondents would accept a process less than 10% more expensive. Also in this case, a hypothetical scenario was presented and respondents were asked to choose between *"A tool that costs 1,000€ and statistically selects 70% men and 30% women, given the same number of male and female applicants"* and *"A tool that costs 3,000€ and statistically selects 60% men and 40% women, given the same number of male and female applicants"*. Results show that 80,2% of respondents preferred to have a three times more expensive process, instead of a more biased one. Then the same question was asked with a different level of bias and a different cost of the process. Respondents had to choose between: *"A tool that costs 1,000€ and statistically selects 60% men and 40% women, given the same number of male and female applicants"* and *"A tool that costs*

5,000€ and statistically selects 50% men and 50% women, given the same number of male and female applicants." In this case, 55,4% of respondents chose to have a five times more expensive, but totally unbiased process. Therefore, out of 81 respondents that gave up on cost efficiency in the first question, 36 considered the level of bias "60% men and 40% women" acceptable and opted for a less expensive process in the second question.

6. Discussion

It is clear that the use of AI systems in recruitment can have numerous benefits, especially the improvement of recruitment quality, time efficiency and cost reduction. Still, there is the risk that AI systems can perpetuate discriminations against certain categories such as women. The present research investigated whether the risk of perpetuating bias towards women is a disincentive to the use of AI in recruitment.

According to the results presented above, some conclusions can be drawn. First of all, it can be said that the idea that selection processes are inevitably subject to unconscious bias is a widespread belief among recruiters. In fact, most of respondents consider bias as something almost always present in recruitment and difficult to remove. That being said, contrary to what one might think, not everyone considers the existence of bias in recruitment a necessarily negative thing. As far as the relationship between bias and AI is concerned, the majority of recruiters believes that AI can help to mitigate bias and improve the quality of recruitment. Although, not everyone completely trusts AI. In particular, the lack of human judgment still seems to be a problem for many. However, even though there is not yet total trust, we can exclude that recruiters see AI as a potential danger for recruitment. On the contrary, it can be said that AI is seen as an opportunity to enhance the recruitment process.

As for the risk of AI perpetuating and exacerbating the human bias, it is possible to state that it constitutes a deterrent for recruiters. In fact, most of respondents would renounce to the improvements in terms of recruitment quality, time efficiency and cost reduction that AI systems could bring, in order to have a less biased process. Nevertheless, it is interesting to note that, for each factor, faced with a reduction in the level of discrimination, many people are more likely to accept the bias. In fact, switching from an AI tool that selects on average 70% men and 30% women, to a tool that selects on average 60% men and 40% women, about 30% of respondents, changed their opinion and accepted the bias. This result allows to say that, although bias does represent a blocker, its entity is an extremely important factor.

Moreover, among the three factors analyzed, the recruitment quality was the most important and the one that fewer recruiters would give up. In fact, half of respondents would accept a biased AI tool if this would lead to an improvement in terms of quality of recruitment. Time efficiency is the second most important factor, while unexpectedly the cost reduction seems to be a renounceable factor for most of respondents.

6.1. Limitations

Considering the research approach, the study presents some limitations that affect the robustness of the results and provide inspiration for future research. First, given the need for experts' opinions, a limited number of responses were collected due to the difficulty of getting in touch with them. Secondly, primary data was gathered through a survey-based questionnaire which entails intrinsic limitations. In fact, the respondents were asked closed-ended questions that inevitably limited the experts' opinions to the ones proposed by the author. This could constitute a source of bias in the research. Moreover, although the questionnaire was anonymous, the subject covered is extremely sensitive and many people may feel pushed towards ethically correct answers. Another aspect to consider is that the scenarios presented in the questionnaire constitute a simplification of reality. When making decisions in a company, there are many more aspects to consider than those proposed, therefore the answers provided may not represent a faithful representation of the reality. Finally, the respondents are not necessarily those who have the responsibility to decide about the implementation of new technologies within their companies. Therefore, the results may not necessarily reflect the companies' point of view. For example, if we look at the factor "cost of the selection process", it was surprisingly considered less important than the others. In fact, most of respondents would be willing to spend more money to have an unbiased selection process. It is likely that, since the cost of the selection process does not directly impact their work, recruiters were more likely say they would sacrifice that metric. Yet, the companies' position may be different.

6.2. Further research

The present research is a theoretical study, based on experts' opinions on hypothetical scenarios. Therefore, the results reflect the attitude of recruiters towards AI, but may not exactly reflect the reality. In order to have a more realistic picture, the study could be integrated with a research that describes the current state of AI in recruitment. First of all, it would be important to investigate how many companies currently use AI in their selection processes and what kind of tools. Then, it would be interesting to conduct a qualitative analysis interviewing the decision-makers of both companies that use and do not use AI. The research should investigate the reason why companies use AI in recruitment, if they implement any measure to take bias under control and if they have any evidence of bias perpetuation. Also, the research should be extended to companies that do not use AI in recruitment to understand the reasons for their choice and to understand to what extent bias is actually a deterrent.

6.3. Post Covid-19 considerations

This research was carried out before the global outbreak of Covid-19, so it is necessary to analyze the validity of the conclusions in a post-pandemic scenario. It is common knowledge that the pandemic has put a strain on the global economy: entire production sectors have had to stop, many workers have lost their jobs and several companies have had to reorganize the way they work.

The recruitment sector has seen an abrupt slowdown due to the decline in hiring, but at the same time, the constraints of the virus have pushed for greater use of technology. Online selection processes have become the usual and this will probably lead to a higher level of implementation of AI systems in the next future (International Labour Organization 2020). In addition, many companies will probably have to reduce their costs, and this could lead to a drop in outsourcing and encourage the use of AI solutions within the company. Even if the diffusion of Covid-19 may not have direct implications on AI bias itself, it is possible that the new companies' needs make it a less relevant problem. For example, the present research shows that most recruiters would be willing to spend more money to have a less biased selection process. But in a post-pandemic context, where many companies will need to be more cost-efficient, will this assertion still be true?

7. Conclusions

The objective of the research was to investigate the problem of AI bias in recruitment. AI applications are becoming increasingly widespread in many areas and recruitment seems to have a great potential to be revolutionized by those technologies. However, despite the numerous benefits, there is the risk that AI may inherit human prejudices and exacerbate discrimination against vulnerable categories, such as women. The title of the research "Just because it's biased, does it mean it's bad?" is a provocative question that expresses the reason that pushed the author towards that topic. The aim was to find out whether the bias represents an impediment to the implementation of AI in recruitment. To do so, the main benefits of the use of AI in recruitment have been considered: hiring a, hiring time and the cost of the hiring process. These three metrics have been opposed to bias in order to understand to what extent recruiters were willing to sacrifice significant benefits to have an unbiased recruitment process. It was possible to conclude that the presence of bias in recruitment is an issue that most recruiters are aware of. Furthermore, the recruiters' perception of AI systems is overall positive, in fact AI is considered a good opportunity to enhance the selection process. However, the existence of bias does represent a disincentive to the implementation of AI in recruitment. That is, recruiters would give up advantages such as improving quality, speeding up the process and reducing the costs in order to have a less biased process. However, low levels of bias are more likely to be tolerated, especially when the trade-off is a decrease in the process quality. Indeed, the quality of recruitment has proven to be the most important factor for recruiters.

It is possible to conclude that the providers of AI tools most likely to succeed are those that demonstrate attention to algorithmic fairness and ensure transparency and equity. Moreover, given the importance of quality, vendors offering AI tools that aim to identify the best profiles, rather than speeding up the selection process or reducing costs by automating tasks, have more chances to be implemented in companies regardless of whether they can introduce bias.

To conclude, this research confirms that the topic of AI in recruitment is still new and there is still no common opinion among recruiters. Therefore, it will be interesting to see how this sector will be impacted in the next future by AI and see how the bias problem will be addressed.

8. References

Ahmed, Owais. 2018. "Artificial intelligence in HR." *International Journal of Research and Analytical Reviews* 971 - 978.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data’s Disparate Impact." *California Law Review* 677-692.

Berry, Michael J. A., and Gordon S. Linoff. 2004. In *Data Mining Techniques: For Marketing, Sales, And Customer Relationship Management*, 263 – 271. Indianapolis: Wiley.

Bogen, Miranda, and Aaron Rieke. 2018. *An Examination of Hiring Algorithms, Equity, and Bias.* Washington: Upturn.

Bostrom, Nick, and Eliezer Yudkowsky. 2014. "The Ethics of Artificial Intelligence." In *The Cambridge Handbook of Artificial Intelligence*, by Keith Frankish and Ramsey William M., 316-334. New York: Cambridge University Press.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 1-15.

Crawford, Kate et al. 2019. *AI Now 2019 Report.* New York: AI Now Institute.

Custers, Bart. 2013. "Data Dilemmas in the Information Society: Introduction and Overview." In *Discrimination and Privacy in the Information Society. Studies in Applied Philosophy, Epistemology and Rational Ethics*, by Bart et al. Custers. Springer.

Dalenberg, David Jacobus. 2018. "Preventing discrimination in the automated targeting of job advertisement." *Computers Law & Security* 615-627.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. 2015. "Automated Experiments on Ad Privacy Settings." *Privacy Enhancing Technologies* 92-112.

Dwork, Cynthia et al. 2012. "Fairness Through Awareness." *Innovations in Theoretical Computer Science Conference* 214- 226.

Entelo Company Website. n.d. *Entelo.* Accessed March 3, 2020. https://www.entelo.com/.

Friedman, Batya, and Helen Nissembaum. 1996. "Bias in computer systems." *ACM Transactions on Information Systems* 330–347.

Hand, David J. 2007. "Principles of Data Mining." *Drug-Safety* 621–622.

Hire by Google. 2019. *How to accurately calculate your cost per hire.* July 25.

HR.com. 2019. *The current and future state of AI in HR.* HR.com.

Ideal Company Website. n.d. *Ideal.* Accessed March 3, 2020. https://ideal.com/.

International Labour Organization. 2020. "COVID-19 impact on the collection of labour market statistics." ILOSTAT. May 11. Accessed May 18, 2020. https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/.

Jobin, Anna, Marcello Ienca, and Effy. 2019 ‘The global landscape of AI ethics guidelines’ Nature Vayena. 2019. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 389-399.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. *Inherent Trade-Offs in the Fair Determination of Risk Scores.* Ithaca, NY, November 17.

Kodiyan, Akhil. 2019. *An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool.* November 18.

Lambrecht, Anja, and Catherine Tucker. 2019. "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads." *Management Science* 2966-2981.

Larson, Jeff et al. 2016. *How We Analyzed the COMPAS Recidivism Algorithm.* New York, May 23.

Lerh, David and Ohm, Paul. 2017. "Playing with the Data: What Legal Scholars Should Learn About Machine Learning." *UCD Davis Law Review* 664 - 695.

Linkedin Talent Solutions. 2018. *Global Recruiting Trends 2018: the four ideas changing how you hire*. Linkedin.

Linkedin Talent Solutions. 2019. *The future of recruiting".* Linkedin.

Liu, Huan, and Hiroshi Motoda. 1998. "Feature Extraction, Construction And Selection." 4 -11. New York: Springer.

Lloyd, Kirsten, and Booz Allen Hamilton. 2018. *Bias Amplification in Artificial Intelligence Systems.* September 20.

Mayson, Sandra Gabriel. 2018. "Bias In, Bias Out." *Yale Law Journal* 2218-1300.

McCarthy, John. 2007. *What is artificial intelligence?* Stanford: Stanford University, November 12.

Mehrabi, Ninareh et al. 2019. *A Survey on Bias and Fairness in Machine Learning.* September 17.

Mya Company Website. n.d. *Mya.* Accessed March 3, 2020. https://www.mya.com/.

Narayanan, Arvind. 2018. *Translation tutorial: 21 fairness definitions and their politics.* Princeton, March 1.

Okolie, Ugo Chuks, and Ikechukwu Emmanuel Irabor. 2017. "E-Recruitment: Practices, Opportunities and Challenges." *European Journal of Business and Management* 116-122.

Oxford University Press. n.d. *Oxford Dictionary.* Accessed March 2020. https://www.oed.com/.

Sanderson, Jon. n.d. "5 Reasons Why AI Can't Replace the Human Touch in Recruitment." *Oliver Parks Consulting.* Accessed March 15, 2020. https://www.oliverparks.com/blog-news/why-ai-cant-replace-recruitment-consultants.

SHRM. 2016. *2016 Human Capital Benchmarking Report.* SHRM.

Silberg, Jake, and James Manyika. 2019. "Notes from the AI frontier: Tackling bias in AI (and in humans)." *McKinsey Global Institute* 1 -6.

Tambiama, Madiega. 2019. *EU guidelines on ethics in artificial intelligence: Context and Implementation.* European Parliamentary Research Service.

Textio Company Website. n.d. *Textio.* Accessed March 3, 2020. https://textio.com/.

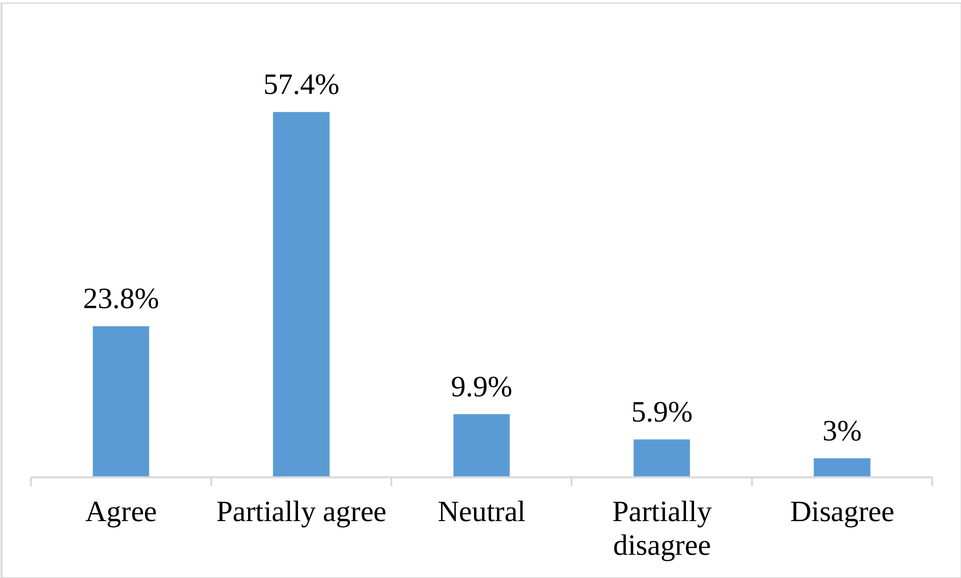
World Economic Forum. 2020. "Global Gender Gap." Insight Report.

Wright, James, and David Atkinson. 2019. *The impact of artificial intelligence within the recruitment.* Carmicheal Fisher Search.

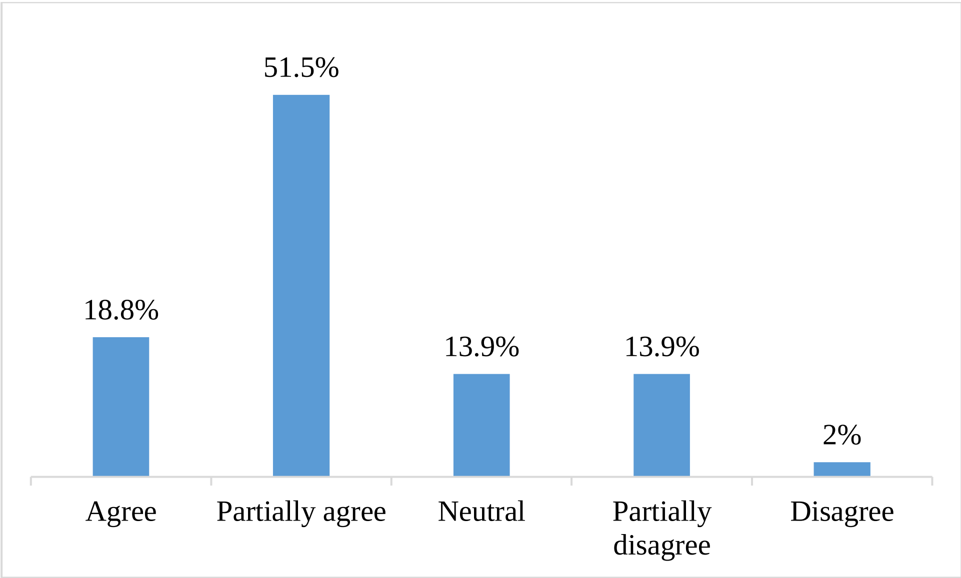
9. Appendix

Appendix 1: Answers from the questionnaire

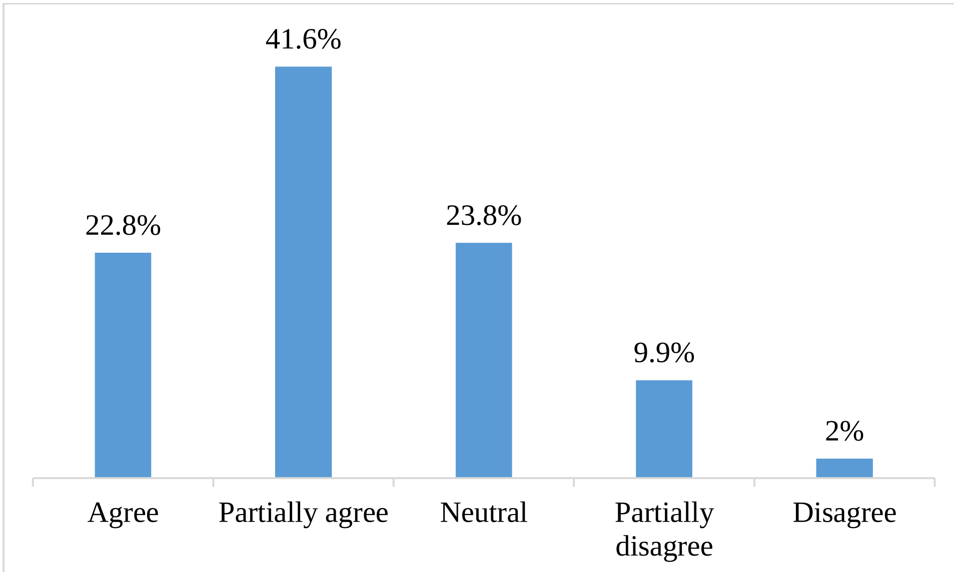
1.1. Hiring processes are always subjected to unconscious bias



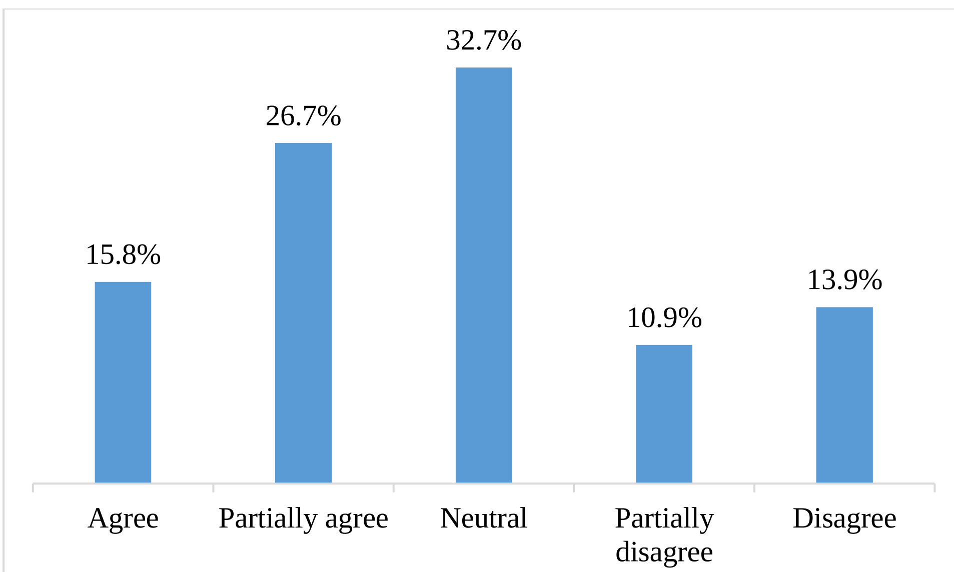
1.2. Bias in hiring processes is hard to remove



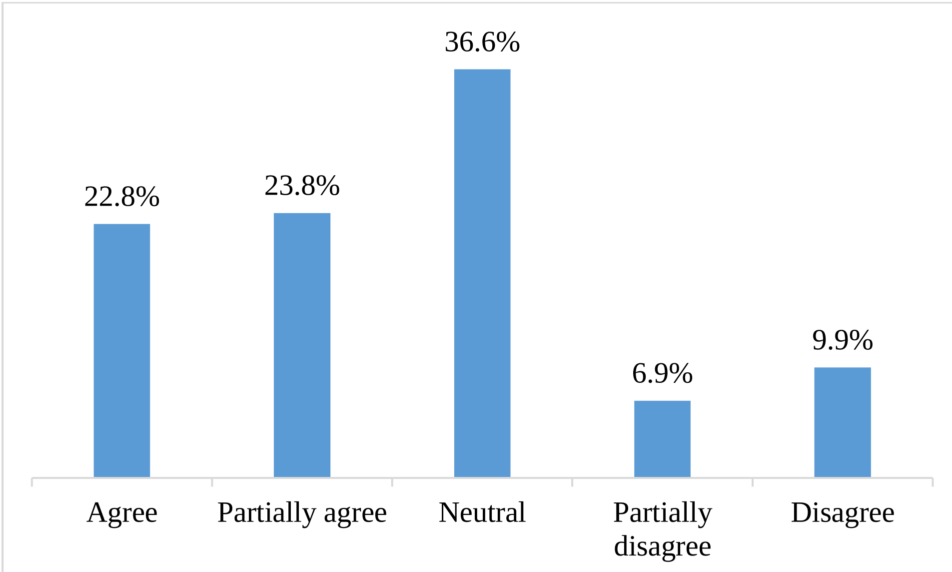
1.3. Bias in hiring processes can be reduced by using technology



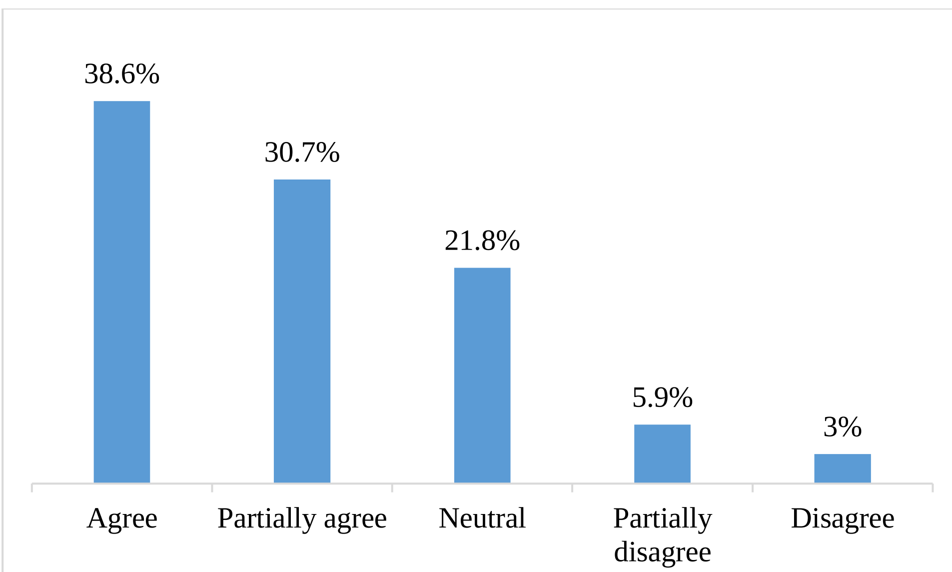
1.4. Bias does not necessarily affect hiring processes in a negative way



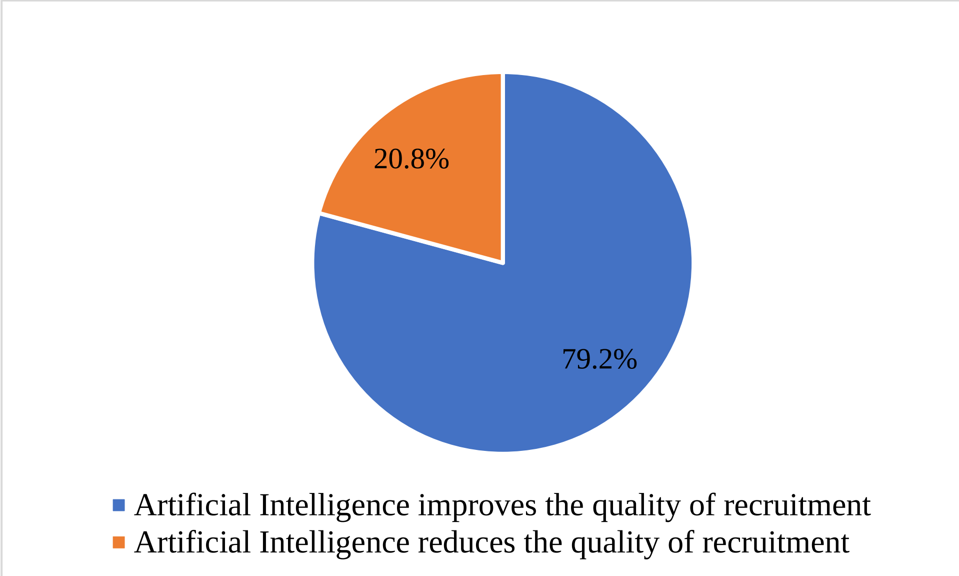
1.5 My company implements measures to reduce bias in hiring processes



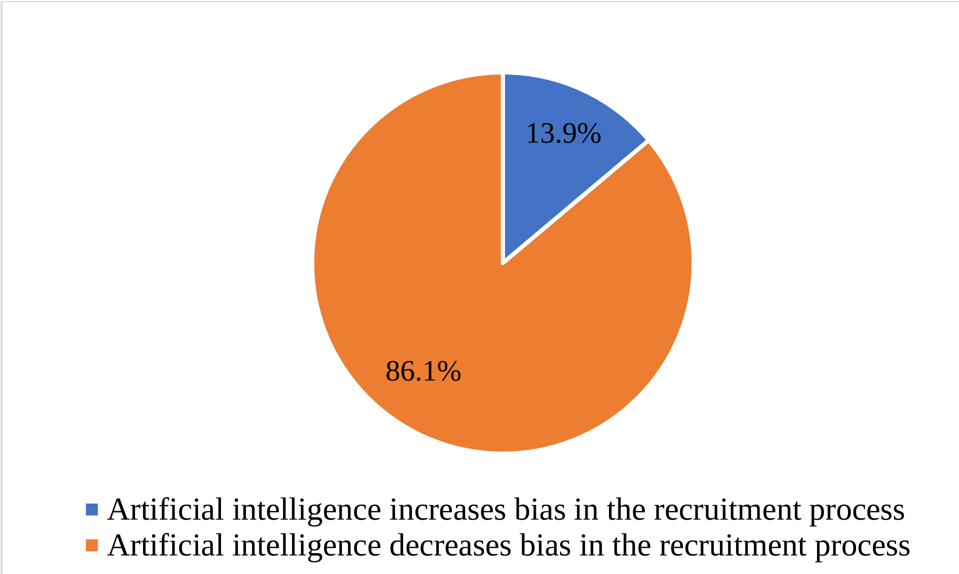
1.6. I'd like my company to put more effort to reduce bias in hiring processes



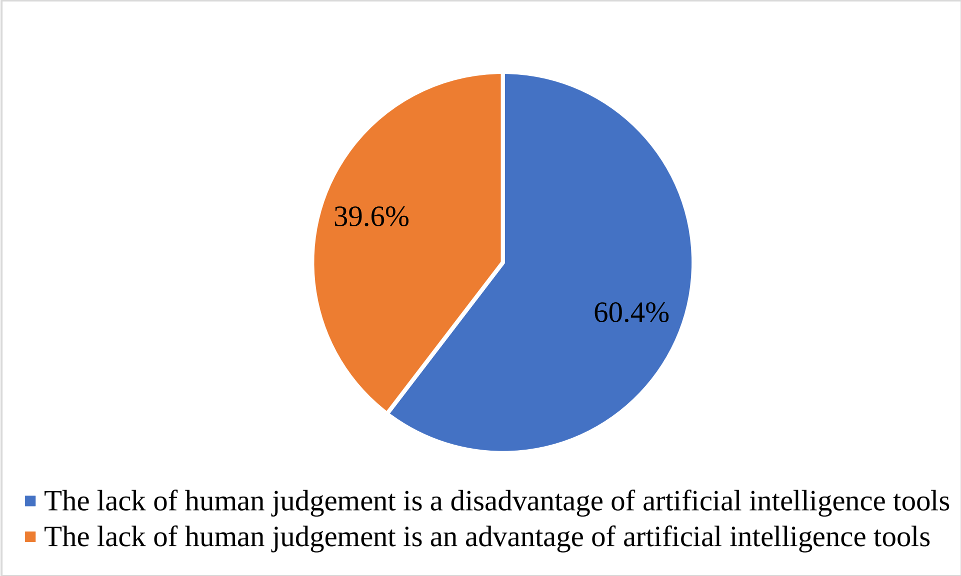
1.7. Pick one alternative between the following



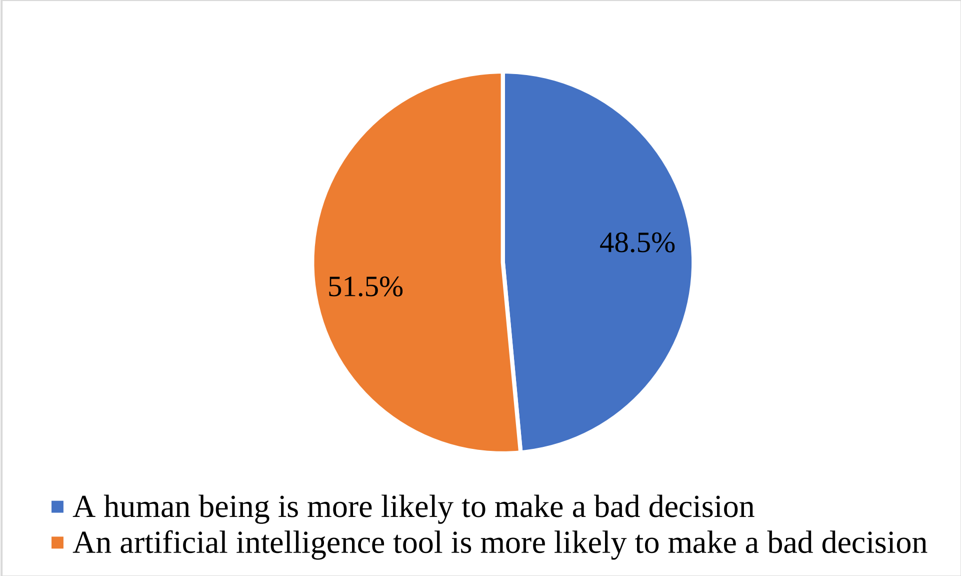
1.8. Pick one alternative between the following



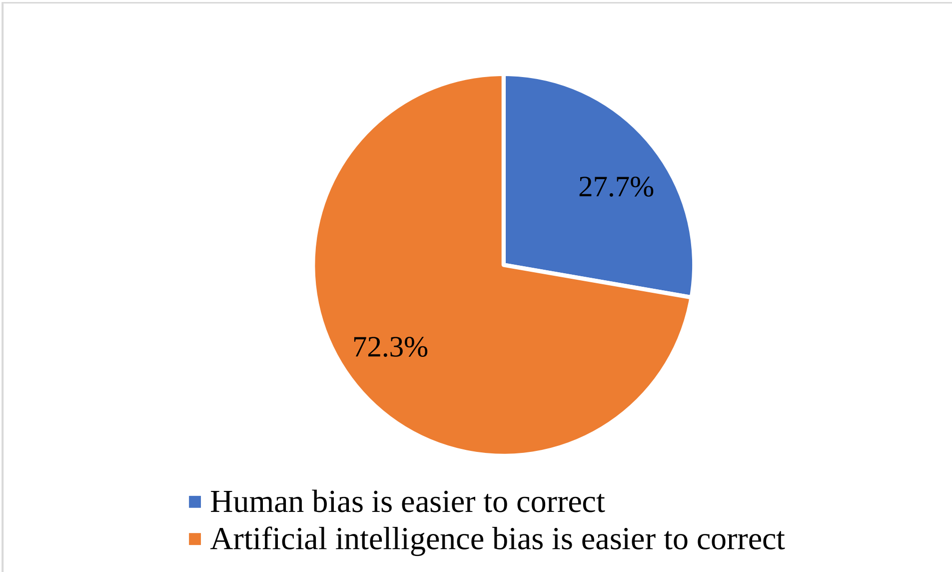
1.9. Pick one alternative between the following



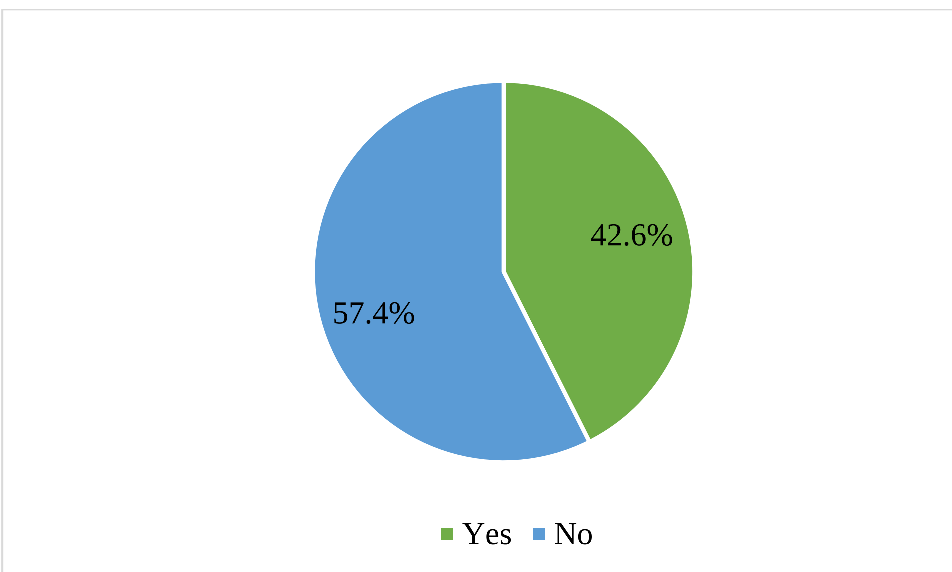
1.10 Pick one alternative between the following



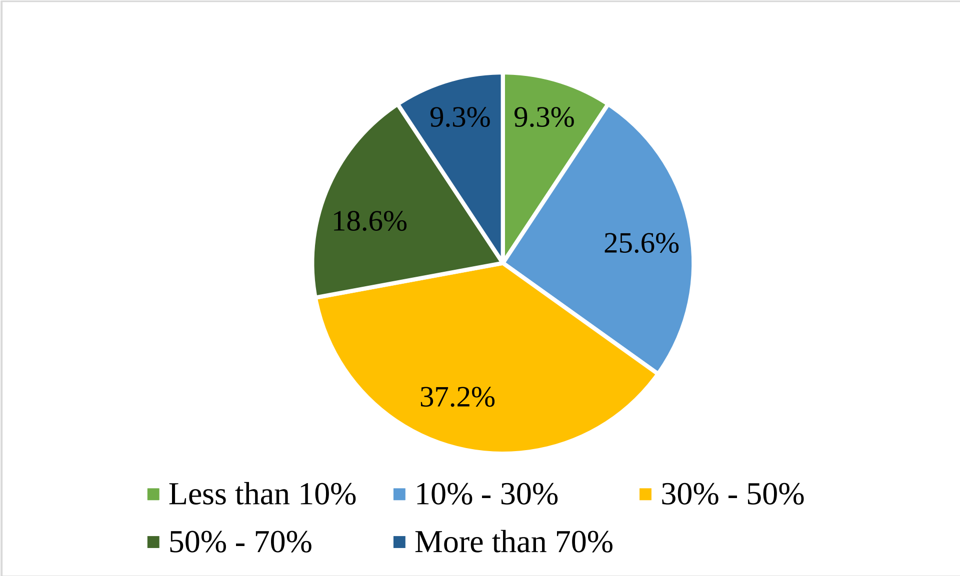
1.11 Pick one alternative between the following



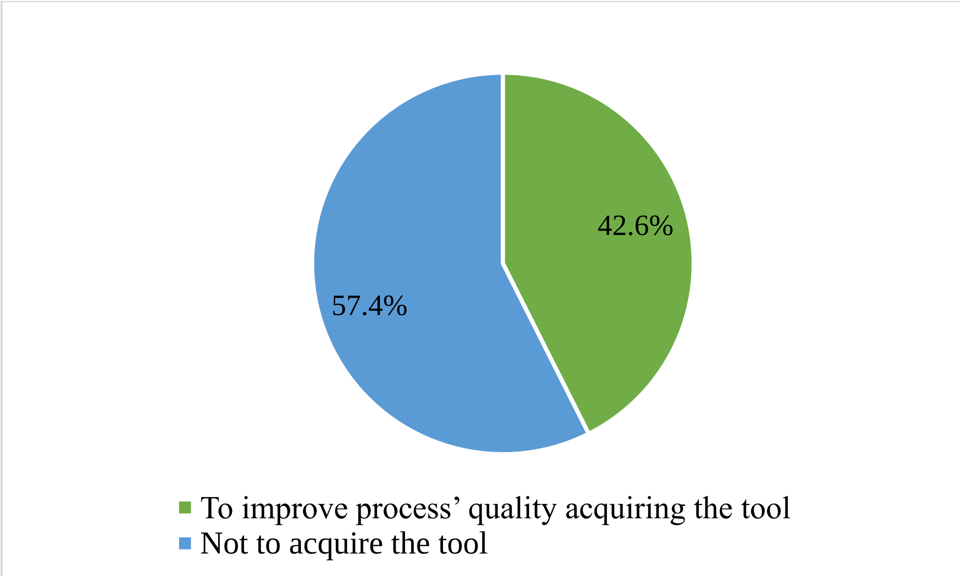
1.12. Would you be willing to give up on the quality of the selection process if this would guarantee the absence of bias?



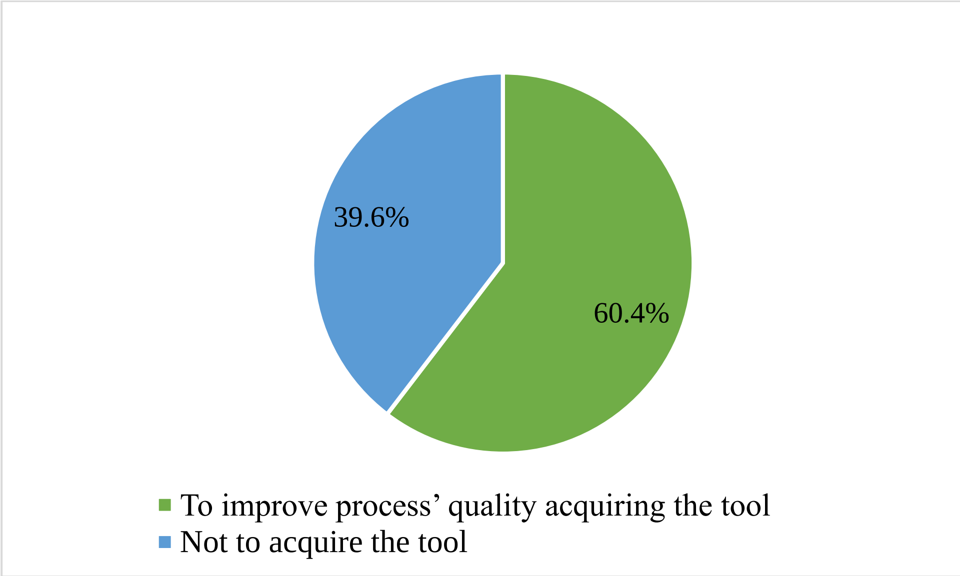
1.13. How much quality (in percentage) would you be willing to give up in order to have a totally unbiased process?



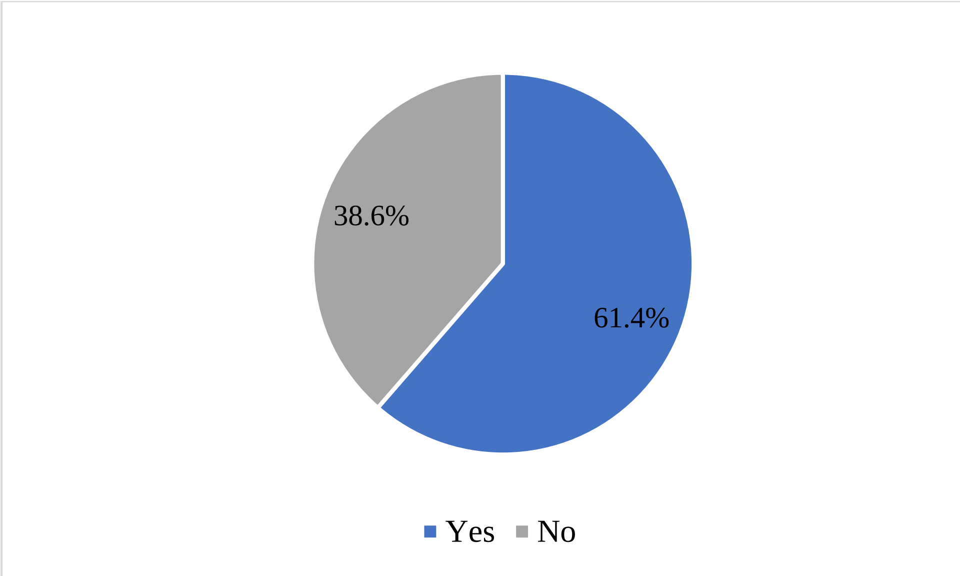
1.14. It is statistically proved that, given the same number of applicants, the tool selects 70% male profiles and 30% female profiles. You would prefer:



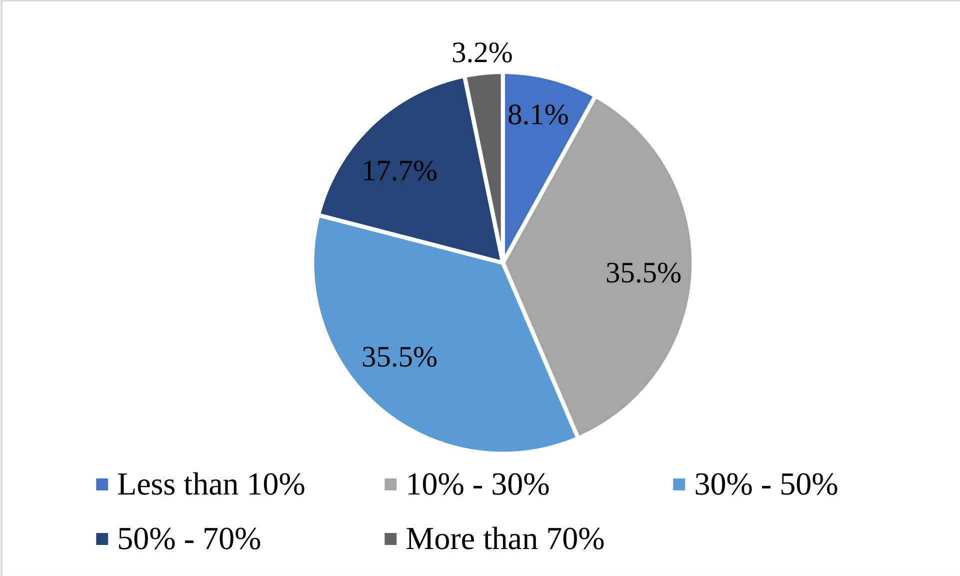
1.15. It is statistically proved that, given the same number of applicants, the tool selects 60% male profiles and 40% female profiles. You would prefer:



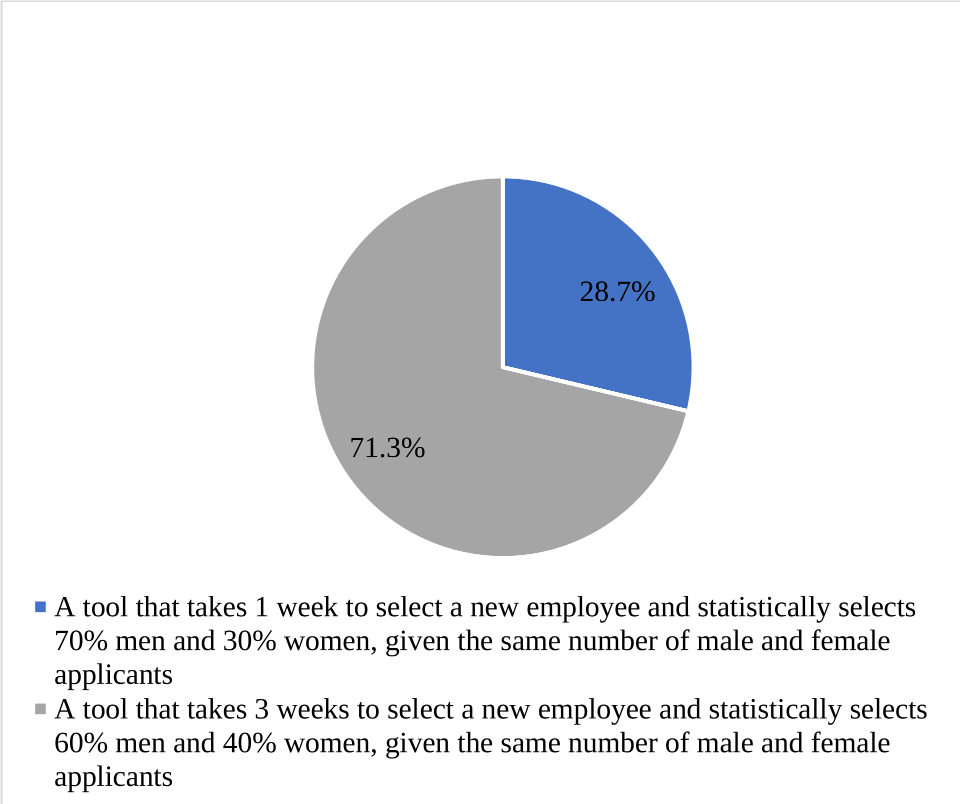
1.16. Would you be willing to give up on the time-efficiency of the selection process if this would guarantee the absence of bias?



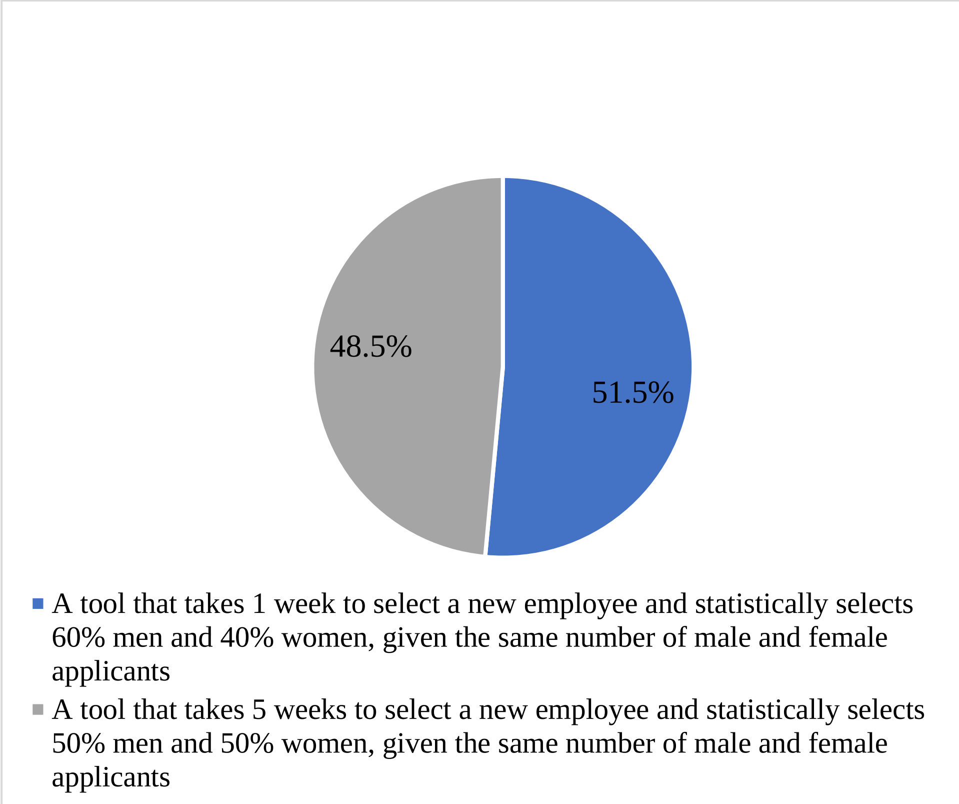
1.17. How much extra time (in percentage) would you be willing to spend to guarantee a totally unbiased process?



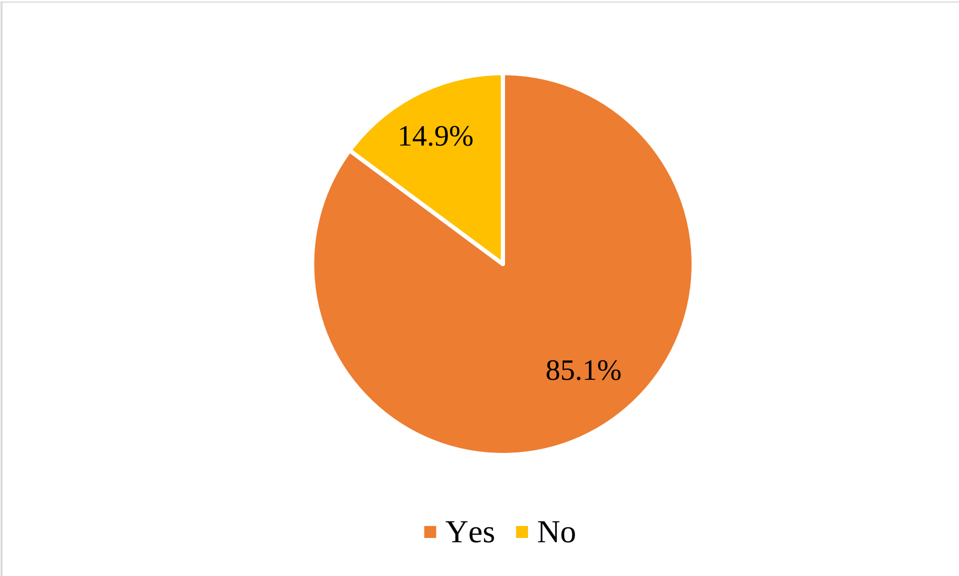
1.18. Everything else being equal, you would prefer:



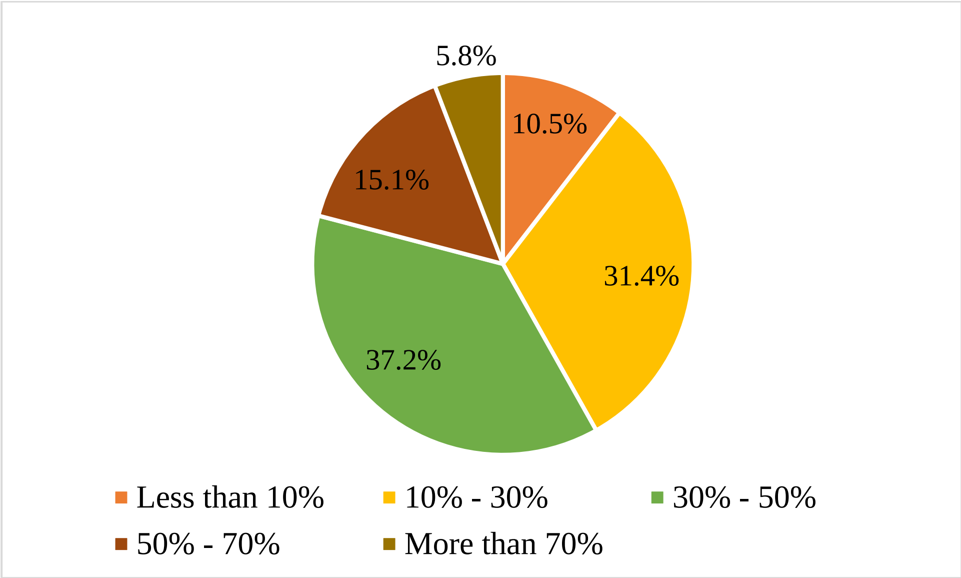
1.19. Everything else being equal, you would prefer:



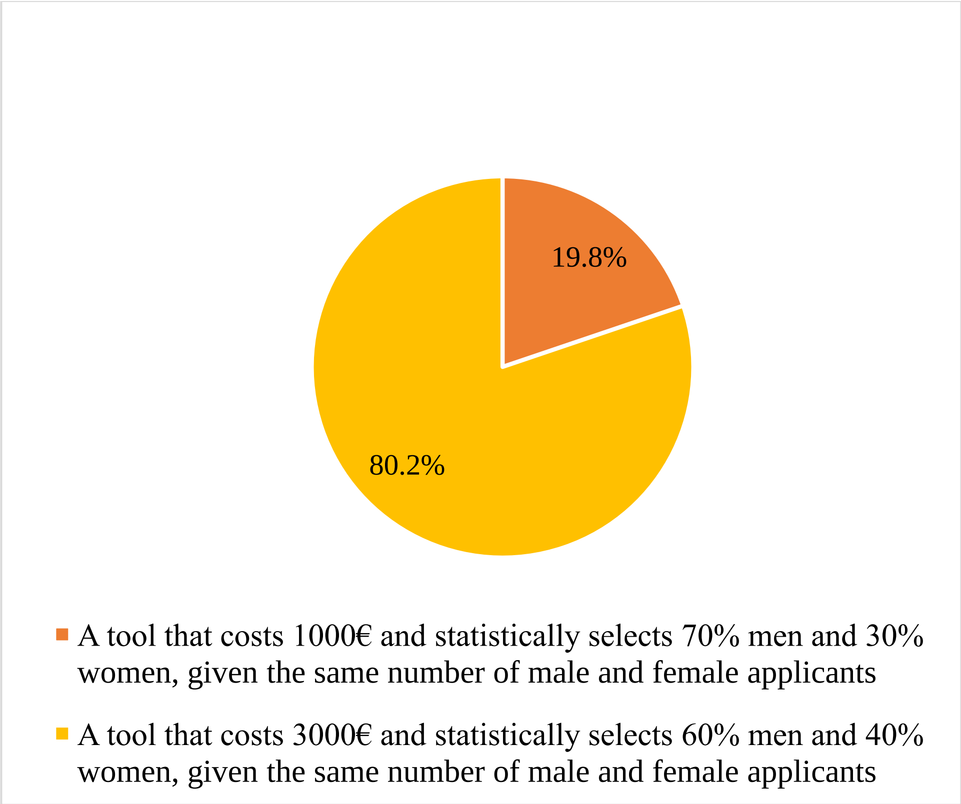
1.20. Would you be willing to pay more for an artificial intelligence tool that could guarantee the absence of bias in the hiring process?



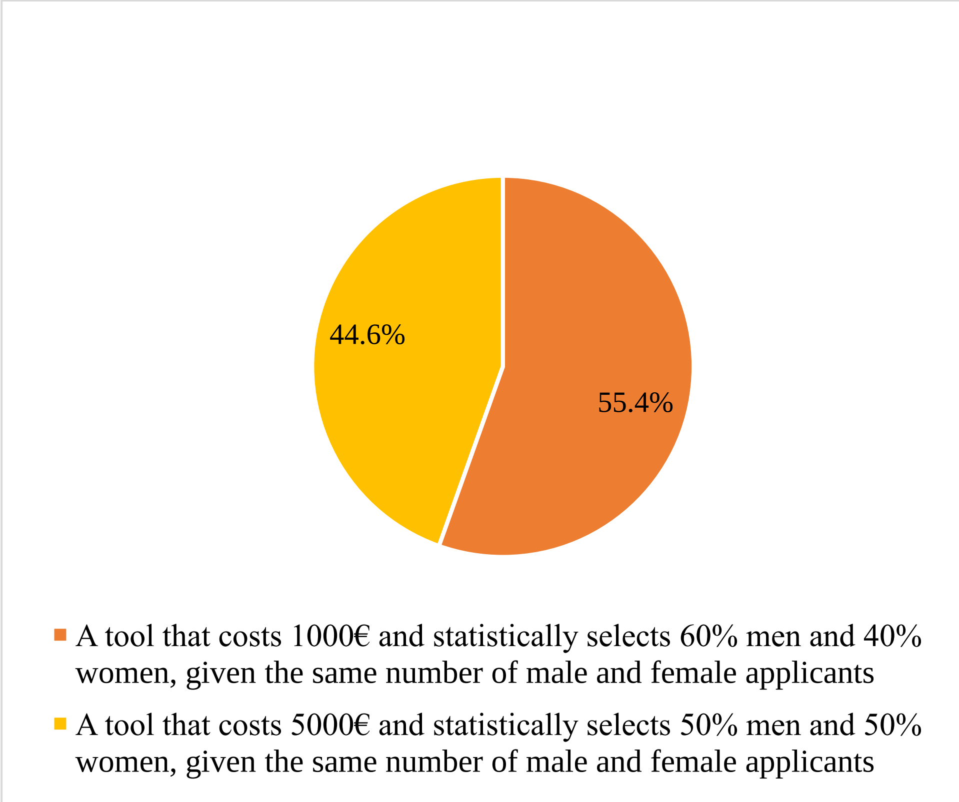
1.21. How much more (in percentage) would you be willing to pay to guarantee non-bias?



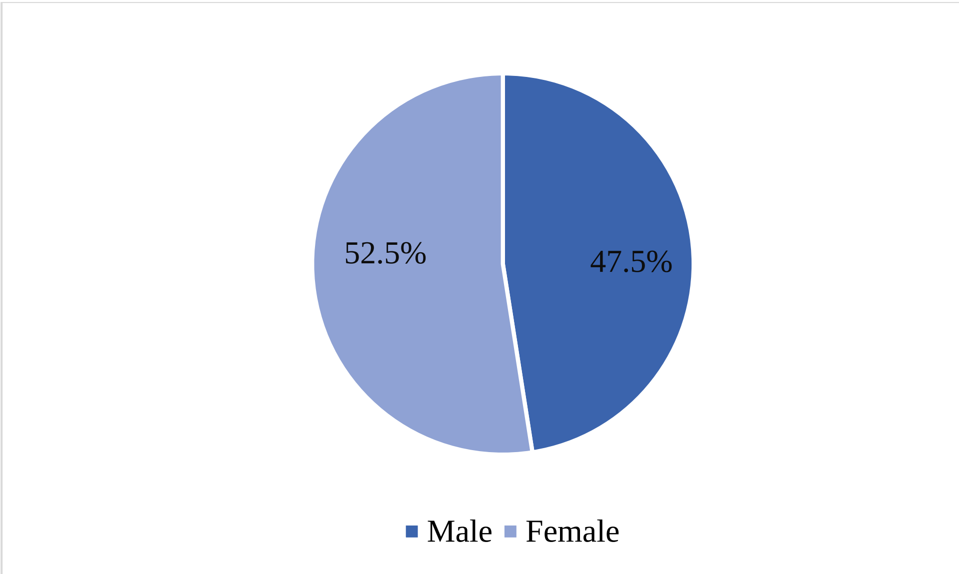
1.22. Everything else being equal, you would prefer:



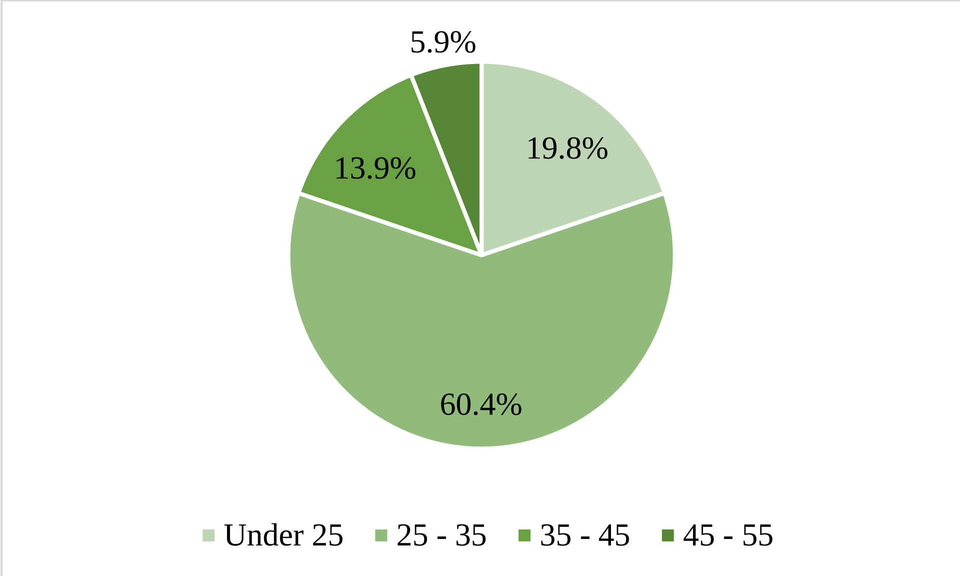
1.23. Everything else being equal, you would prefer:



1.24. Gender



1.25. Age



1.26. Country

Brazil	1%
Cambodia	1%
France	1%
Germany	10.9%
India	1%
Italy	40.6%
Japan	1%
Lithuania	1%
Mexico	1%
Portugal	18.8%
Romania	1%
Saudi Arabia	1%
Spain	10.9%
Switzerland	2%
United Arab Emirates	2%
United Kingdom	3%
USA	3%

1.27. Industry

Apparel/Fashion	1%
Automotive	1%
Banking	5.9%
Consulting	28.7%
Consumer Goods	7.9%
Education	4%
Entertainment	1%
Financial Services	4%
Food/Beverages	1%
Health Care	7.9%
Hospitality	4%
Information Technology	11.9%
Insurance	1%
Internet	4%
Legal Services	4%
Luxury Goods	4%
Manufacturing	3%
Real Estate	1%
Retail	3%
Telecommunications	1%
Transportation	1%