TU WIEN Informatics

Systematischen Bekämpfung unbewusster Vorurteile in Vorstellungsgesprächen mit Anonymisierungstechnologie

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medieninformatik und Visual Computing

eingereicht von

Lukas Witzany, BSc.
Matrikelnummer 00527644

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioğlu
Mitwirkung: Dipl.-Ing. Dr. techn. Florian Cech

Wien, 7. Mai 2024

Lukas Witzany	Hilda Tellioğlu

TU WIEN Informatics

Systematically Addressing Unconscious Bias in Job Interviews with Anonymization Technology

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics and Visual Computing

by

Lukas Witzany, BSc.
Registration Number 00527644

to the Faculty of Informatics

at the TU Wien

Advisor: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioğlu
Assistance: Dipl.-Ing. Dr. techn. Florian Cech

Vienna, 7th May, 2024

Lukas Witzany	Hilda Tellioğlu

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Lukas Witzany, BSc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 7. Mai 2024

Lukas Witzany

Acknowledgements

Completing this work was anything but a small feat alongside juggling a job and a family. Consequently, it took over four years amidst all the obstacles life had in store. Small wonder motivation got lost a couple of times along the way.

My family, however, always helped me pick up again right where I had left off and helped me eventually follow through. I especially want to thank my wife Lena for her love and support throughout and my daughters Lilja and Lotte for providing levity in times when there was seemingly no progress to be made. I also thank my parents, my mother specifically, for giving me time to write by taking care of the kids and helping out wherever possible.

As far as the more immediate contributions to this thesis, I want to thank Hilda Tellioğlu for the opportunity to work on this topic and the continuous support along the way. To Florian Cech, who contributed greatly to this work by supporting me and facilitating the project in all its stages, I owe special gratitude. I also want to thank the participants of the design workshop for contributing their time and their valuable input. In this spirit, I also want to acknowledge the contribution of Hayley Green who helped conduct participant interviews in the first event.

The people at the *TU Career Center*, especially Michaela Unger and Christina Habetler, should not go unrecognized either. Without their initiative and their coordination efforts evaluating the DEBIAS system would not have been possible in this form.

Lastly, I want to thank everyone who participated in the two evaluation events and all consecutive applications of the DEBIAS process since.

Kurzfassung

Die gängige Praxis in Bewerbungsverfahren ist nachweislich anfällig für kognitive Verzerrungen und unbewusste Vorurteile. Dabei sind Faire Auswahlverfahren nicht nur ein moralisches Gebot – Vielfalt am Arbeitsplatz ist ein wesentlicher Treiber für Innovation und Unternehmenserfolg. Die vorliegende Arbeit widmet sich der Konzipierung und Implementierung eines soziotechnischen Systems zur Durchführung anonymer strukturierter textbasierter Bewerbungsgespräche. Dabei wurde untersucht wie groß das Potenzial eines text-basierten Jobinterviews ist Biases zu beseitigen und wie sich ein solches Verfahren auf die Fähigkeit der Bewerber:innen auswirkt sich selbst zu präsentieren. Schließlich wurde beleuchtet welchen Einfluss der Ansatz auf die Fähigkeit der Interviewer:innen hat, sich einen Eindruck von den Bewerber:innen zu verschaffen. Mehrere Firmen interviewten dabei Studierende der *TU Wien* in einem zweistufigen Prozess, bestehend aus einem völlig anonymisierten strukturierten Chat-Interview gefolgt von einem offenen Gespräch von Angesicht zu Angesicht. Die beiden Iterationen des Systems wurden jeweils im Rahmen eines vom *TU Career Center* veranstalteten Recruiting-Events evaluiert. Dabei wurden qualitative Daten in Form von Interviews mit Studierenden und Unternehmensvertretern gesammelt, welche transkribiert und mittels qualitativer Inhaltsanalyse ausgewertet wurden. Quantitative Daten wurden im Rahmen der Erhebung von Ranglisten der Bewerber:innen durch die Unternehmen gewonnen. Diese Daten wurden statistisch ausgewertet, um die Validität des Ansatzes in Bezug auf die Bewerber:innenauswahl zu beurteilen. Das entwickelte System "DEBIAS" zeigte großes Potenzial Vorurteile im Vorstellungsgespräch zu reduzieren und ein faireres Interviewverfahren zu ermöglichen. Die Fähigkeit der Bewerber:innen sich den Firmen gegenüber selbst zu präsentieren wurde in der Regel nicht merklich beeinträchtigt. Im Gegenteil: der zweiteilige Interviewprozess führte oft zu verringerter Nervosität und könnte so sogar einen Vorteil für die Bewerber:innen darstellen. Auch Firmenvertreter:innen gaben an die Kandidat:innen in der Regel ausreichend anhand des Chatinterviews beurteilen zu können. Die Kombination aus anonymen Textinterview und offenem Interview zeigte sich als verlässlicher Indikator für die Kandidat:innenauswahl mit einer starken Korrelation für alle Firmen in der zweiten Iteration. Diese Forschungsarbeit leistet einen wertvollen Beitrag zu den Bemühungen um die Schaffung eines faireren und effektiveren Rekrutierungsprozesses. Die Ergebnisse zeigen, dass anonymisierte, strukturierte Textinterviews als vielversprechendes Instrument zur Minderung von Vorurteilen eingesetzt werden können und dabei gleichzeitig das Potenzial haben eine positivere Bewerbungserfahrung für Bewerber:innen zu schaffen.

Abstract

Hiring practices are demonstrably prone to bias, leading to the exclusion of qualified candidates and hindering the establishment of diverse workplaces. While workplace diversity is not only a moral imperative but also a critical factor in organizational success, blind hiring techniques, despite their effectiveness in mitigating bias, are rarely put into practice. This research investigates the potential of anonymized, structured text-based job interviews to reduce bias in the hiring process. A novel sociotechnical system, DEBIAS, was developed and evaluated in two real-world recruiting events organized by the *TU Career Center* with the aim of investigating the potential of text-based job interviews to eliminate bias and assess their impact on applicants' self-presentation and interviewers' ability to form adequate impressions of candidates. In two iterations, companies interviewed students of *TU Wien* utilizing a two-step process: a completely anonymous structured chat interview followed by an open face-to-face conversation. Qualitative data in the form of interviews with students and company representatives was collected, transcribed, and analyzed using qualitative content analysis. Quantitative data were collected by asking companies to rank the applicants. These data were statistically analyzed to assess the validity of the approach for applicant selection. The developed system showed great potential to reduce bias in job interviews and enable a fairer interview process. Applicants' ability to present themselves to companies was generally not impaired. On the contrary, the two-part interview process often led to reduced nervousness and could even present an advantage to applicants. Company representatives also indicated that they could generally assess candidates adequately based on the chat interview. The combination of anonymous text interview and face-to-face interview proved to be a reliable indicator for candidate selection with a strong correlation for all companies in the second iteration. This research makes a valuable contribution to the efforts of creating a fairer and more effective recruitment process. The results show that anonymous, structured text interviews present a promising tool to reduce bias and at the same time have the potential to create a more positive recruiting experience for applicants.

Contents

CHAPTER 1

Introduction

The demographic makeup of Europe's population is changing and with it the labour market. Ongoing trends such as women making up an ever-increasing portion of the available workforce and an overall population that is continually aging challenge common hiring practices [1, 2]. Due to declining birth rates, ongoing migration in combination with a higher fertility rate among the migrant population will lead to an ever-increasing share of the population having a migration background [3]. The *Austrian Economic Chambers* acknowledges these trends, stating that it will become increasingly necessary to make use of previously underrepresented parts of the labor force potential [4]:

> *"Additionally, it is going to be necessary to increasingly make use of the complete labor force potential. This especially applies to women, older strata of the population, and people with migration background."* [1]

This statement is evidence of a growing awareness that workplace diversity is becoming not only a moral implication and a question of equity, but also a detrimental factor to the thriving of a business and even a necessity in the light of ongoing demographic change.

When it comes to the hiring process, however, women, migrants, and people aged 55 and older have traditionally been subject to discrimination [5]. This means that many companies not only fail to grant a fair job application process to candidates, but they also miss out on an important part of the available workforce. In doing so, companies hurt their economic potential. Not only has workplace diversity been shown to foster innovation and create a more pleasant environment to work in, but it also seems to be a veritable success factor for an organization's economic development [6, 7, 8, 9].

[1] Translated from German by the author. Original quote published by the *Austrian Economic Chambers*: "Darüber hinaus wird es in Zukunft notwendig sein, das gesamte Erwerbspotenzial verstärkt zu nutzen. Das betrifft insbesondere Frauen, ältere Bevölkerungsschichten und Personen mit Migrationshintergrund."

Oftentimes, what lies at the bottom of hiring discrimination are unconscious biases, a phenomenon that, often unbeknownst to the person affected by them, can impact almost every aspect of our day-to-day experiences. While the negative impact of overt biases might be apparent, unknowingly held biases can lead to consequences that are even more perfidious. These biases are deeply ingrained in our brains and affect the way we view the people we encounter in our daily lives. How we judge other people is shaped by a combination of our past experiences and the specific ways in which human perception operates.

Hiring employees can, like most decision-making processes, be especially prone to biases. This includes the evaluation of candidates' resumes as well as the whole interviewing process that job applicants are subjected to. Many companies try to address this issue with diversity or unconscious bias training, a method with debatable effectiveness. A 2016 meta-analysis reviewed 40 years' worth of data on unconscious bias training and found that the effect of the particular training efforts varied greatly depending on the methods employed [10]. While these sorts of interventions can help to mitigate biases, they hardly eliminate them. The sheer knowledge of a bias is not enough for a person to be able to prevent its influence.

Due to the nature of bias being largely out of a person's control, other approaches appear to be more promising. Anonymization techniques have successfully been used in recruitment scenarios to level the playing field for applicants. A prominent example of this is how the chances for female musicians to make it into a symphony orchestra increase significantly by concealing the applicants' identities through blind auditions [11]. An overall positive impact of anonymous applications was found concerning discrimination based on gender and ethnicity [12]. However, anonymous job applications are far from being widely adopted, as they are often regarded as more of a nuisance than an asset by employers. This becomes apparent when examining a position paper published on the topic of anonymization in the application process by the *Austrian Economic Chambers*, focusing entirely on potential downsides like additional costs, all the while disregarding possible benefits of such measures outright [13].

1.1 Aim of the Work

This work aims to examine the viability of anonymization efforts in the job interview portion of the hiring process by developing a prototypical sociotechnical system designed to mitigate known biases. This will involve the creation and implementation of a software solution that facilitates anonymous job interviews scheduled to occur at two recruiting events hosted by the *TU Career Center*, providing an opportunity for real-world testing and evaluation of the system. Compared to the systems currently available, this work is focused on the job interview stage of the recruitment process. While the few available tools to provide ways to conduct anonymous technical interviews are limited to a small set of professions (cf. Section 3.5), this work seeks to address job interviews as they are prevalent in a variety of different industries to answer the following research questions:

RQ1 *Minimizing Bias in Interviews: What are the main requirements for a job interview process optimized to reduce bias while maintaining buy-in from decision-makers?*

RQ2 *Balancing Anonymity and Selection Needs: To what capacity can anonymized interview processes address bias while still meeting hiring managers' information needs?*

RQ3 *Predictive Power of Anonymized Interviews: How reliable are anonymous structured job interviews in predicting hiring decisions?*

RQ4 *Text Interviews and Self-Presentation: Do text-based interview formats impact applicants' ability to showcase their personality and skills compared to traditional formats?*

1.2 Methodological Approach

The development process will involve the identification of key requirements based on a review of existing literature on implicit/unconscious bias and potential mitigation strategies. A pre-study featuring interviews with company representatives involved in or responsible for job application processes will be conducted to gather information on the status quo regarding the awareness of unconscious biases and currently employed mitigation strategies to address issues of diversity and discrimination specifically in the hiring process. The information gathered will inform the system design regarding the typical information needs of people conducting job interviews. The interviews will be evaluated using qualitative content analysis as proposed by Mayring [14] to identify common themes and possible parallels as well as differences among participating companies. In conjunction with the findings from the literature research, the stakeholder interviews are going to help answer research question RQ1 by serving as the basis for the requirements analysis. Based on the identified characteristics, a system featuring a structured anonymous process for conducting job interviews will be designed. Employing a human-centered design approach, the data gathered in the pre-study will function as a jumping-off point for a design workshop held with HCI and CSCW scholars and UI experts, which will yield the cornerstones of the system's design, thus answering research question RQ2.

The system prototype will be implemented as a web application based on the Python Pyramid Framework using a *Docker* compose setup for streamlined development and easy deployment. The open text messaging protocol XMPP will serve as the basis for the textual exchange. The system will be employed in two recruiting events, hosting 7 different companies, where it will be used to interview potential candidates for internship positions. Exit interviews conducted with job applicants as well as the participating company representatives will provide insights into their respective experiences. The gathered data will be analyzed using qualitative content analysis according to Mayring [14] to answer research question RQ3.

Additionally, recruiters will evaluate the applicants' performances in the anonymous interviews for each question. This will result in an applicant ranking based on the per-question performance. The recruiters will be asked to provide a ranking of their own at the end of the process, which does not necessarily match the per-question ranking. To answer research question RQ4, the validity of the per-question ranking for hiring decisions will be reviewed by means of statistical evaluation using the mean rank deviation as well as the Spearman rank correlation coefficient [15] comparing the calculated ranking and the manually determined ranking to assess if the criteria to make a hiring decision are in accordance with the evaluations of applicant performance on a per-question level.

1.3 Structure of the Work

The work is based on a four-part structure. Part 1 is concerned with the theoretical background on implicit biases in Chapter 2 and the promotion of workplace diversity in Chapter 3 and provides the foundation for the design and implementation efforts covered in Part 2 and Part 3 respectively.

Based on the findings resulting form the pre-study described in Chapter 4, the design process employed to arrive at the initial prototype is outlined in Chapter 5.

Part 3 is concerned with the details on the implementation and evaluation of the system with the first iteration being covered in Chapter 6 and the second iteration in Chapter 7.

At last, in Chapter 8 of Part 4, I provide an analysis of the findings presented in Part 3, discussing the most relevant results, as well as their limitations and implications for future research.

Part 1

Theoretical Background

CHAPTER 2

Implicit Prejudice

The prevalent understanding of what prejudices are and how they function is based on views that have been formulated generations ago and are still perpetuated in mass media to this day. Prejudices are viewed as being primarily rooted in personal animus and hatred. Consequently, the general understanding still is that these negative attitudes towards social groups are best addressed on an individual level [16]. The Academy Award-winning 2018 movie *Green Book* about a black musician and his heavily prejudiced white driver is one fairly recent example of the perpetuation of the notion racial prejudice is primarily an individual problem that is best addressed by changing individual hearts and minds [17]. Today's scientific understanding of prejudice and stereotyping, however, finds that prejudices do not necessarily require underlying personal aversion or hostility. Rather, they are often *implicit* so that the person acting based on prejudice does so unwittingly, unintentionally, and often without even being aware of the phenomena at play [16]. Human cognition is a complex process that has been found to function largely out of a person's conscious attentional focus. Many mental processes work implicitly [18]. This means, for instance, that a person making a decision is practically never fully aware of how they arrived at the said decision (even when they think they are) because the decision-making process is guided by implicit mechanisms that are out of the person's control and even function outside the bounds of the person's consciousness.

Research suggests that the mere awareness of a stereotype about a social group has an influence on an individual's social judgment regarding a member of said group, regardless of whether the individual subscribes to the stereotype or not. This has repeatedly been reproduced in experiments, demonstrating that how we evaluate a person is implicitly influenced by the recent exposure to judgment-relevant information [16]. An influential 1989 study [19] demonstrated that this applies to subjects who expose explicit prejudices as well as low-prejudice individuals. Both groups were found to be equally knowledgeable of a specific cultural stereotype. Said stereotype was then automatically activated in the presence of a member of the stereotyped group for both high-prejudice and low-

prejudice individuals. Social stereotypes are learned from an early age and were measured to be present even in children [20]. Research suggests that implicit attitudes about social categories emerge at a young age. Children as young as six showed an implicit association of the social category *White* with a positively connoted concept (good) while simultaneously associating *Black* with a negatively connoted concept (bad).

The study used a widely adopted method called the *implicit association test* (IAT) [21], a computer-based technique, developed to measure the strength of a person's implicit associations between certain mental concepts in memory. It typically requires the test subjects to rapidly assign a specific attribute to one of two target concepts by pressing specific buttons. Response times are recorded and interpreted such that faster responses indicate a stronger in-memory association between the concepts. While the term *implicit* is frequently used interchangeably with the term *unconscious*, a more clear-cut definition of *implicit* limits the meaning to what is measured indirectly, for instance, using the IAT or a similar method, as opposed to what is measured directly (i.e. self-reported data) [22]. While the IAT is not without its critics, it marked a paradigm shift for the field of social psychology which had relied on individual self-reports in understanding human behavior for the longest time, assuming that thoughts and feelings were explicitly accessible and endorsed by the individual [23].

It is important to note that implicit prejudice characteristically works ubiquitously and usually goes unnoticed. It is involved in most of our day-to-day information processing without our explicit awareness and requires no personal animus to become active. In fact, implicit prejudice is rather immune to deliberate attempts at avoiding individual biases, as has been demonstrated in experiments where stereotype salience was manipulated in unobtrusive or even subliminal ways, without the participant's knowledge [16].

2.1 How Bias Translates Into Discrimination

Consequently, implicit prejudices also affect our professional lives where numerous studies demonstrate the widespread discriminatory effects of implicit prejudice. Several field experiments found evidence of ethnic discrimination in the hiring process. A foreign-sounding name on the application alone severely diminishes the chances of being invited for a job interview [24, 25].

In a study conducted by the US National Bureau of Economic Research, researchers employed a standardized method known as *correspondence test* to find that applicants with names typically more common in the black population ('Lakisha' and 'Jamal') received up to 50% fewer responses than applicants with typically white names ('Emily' and 'Greg') when they applied for the same jobs. A 2016 meta-analysis of 25 years worth of field experiments conducted across OECD countries found similar phenomena across countries and economic situations, demonstrating the prevalence of discrimination patterns even in the presence of anti-discrimination legislation. Social minority candidates generally need to send out significantly more job applications than their majority equivalents to be called back for a job interview [26]. A comparable correspondence testing study conducted

by the Institute for Advanced Studies (IHS) demonstrated that similar results could be reproduced in Austria [27]. Applications by identities with a migration background received significantly lower response rates than applications by 'Austrians' even though the fictitious applications explicitly mentioned that the applicants were Austrian citizens and had received an education in Austria. The study also explored response rates separated by identity and found that while all migrant identities showed a response rate lower than Austrian, the worst response rate was found for Nigerian applicants, who had to send out almost twice the amount of applications as compared to Austrians, to be invited for an interview. The same study also found significant differences in wages between citizens and non-citizens based on a statistical analysis of the available labor market and micro-census data. Adjusted for factors such as economic sector, job, and job position, a remaining gap of 2.5 percentage points for men and 3.2 percentage points for women was attributed to discrimination against non-citizens.

Nationality or ethnic background are hardly the only factors playing a role in labor market discrimination. Correspondence studies have been conducted to evaluate the influence of a host of different factors, such as gender, sexual identity, sexual orientation, age, religion, and disability status on an applicant's chances to receive a response to their application, not only for hiring but also for fields such as the housing market. An overwhelming majority of studies reported negative treatment effects effectively translating into discrimination against minorities [28]. The role of gender-based discrimination has been well documented, particularly in the academic sphere. In a 1999 study, Steinpreis et al. found that not only were both male and female participants more likely to hire a male job applicant than a female applicant with equal qualifications but they also tended to evaluate the professional achievements of male candidates more positively [29]. These results are consistent with earlier findings showing that the heads of departments were significantly more likely to offer better job entry levels to male candidates than female candidates. Research demonstrating that equally qualified female undergraduate candidates were evaluated to be less worthy of being hired than their male equivalents offers further evidence for the existence of a gender bias [30]. In the experiment, faculty staff rated male applicants as more competent and also selected a higher starting salary for male applicants. Though women do appear to bear the brunt of discriminatory effects, implicit gender biases do not exclusively affect women. For female-dominated occupations (meaning occupations made up of an employee base consisting of 80% women or more) a bias towards favoring female applicants was found [31]. Male applicants were significantly less likely to receive a callback. A fact that can potentially be attributed to gender stereotyping. In jobs that are perceived as being more appropriate for women, male applicants may receive less favorable evaluations because they do not fit the common image of the prototypical representative of the occupation.

While a person's name can already give away information on their gender or a possible migration background, yet another factor playing a role in hiring decisions takes effect only once the visible appearance of an applicant is revealed. The so-called *beauty premium* describes a phenomenon leading to people exhibiting higher physical attractiveness being

generally paid higher wages. In their 1994 study, Hamermesh and Biddle examined three distinct sets of real-world household data to explore the role of a person's physical attractiveness on their wage [32]. They found that all other things equal, people with below-average attractiveness generally were paid less than average-looking individuals. They even identified a premium in wages for good-looking people, though the penalty for individuals with below-average looks was found larger in effect than the premium received by individuals with above-average looks. A trend towards a wage penalization of below-average-looking individuals was found across countries with different effect sizes [33]. An Australian study investigating the potential role of self-perception and confidence on individual wages found that differences in wages could not be explained by individual confidence levels but were rather a direct result of an individual's external perception [34].

The *Beauty Premium* can be explained by a cognitive bias found in human perception known as the *halo effect*. This phenomenon describes the tendency for a positive impression concerning one aspect of a person or even an object, to positively influence one's opinion in other areas [35]. One can easily see how this can become a problem in the context of the hiring process. The first impression we get of a person might all but determine how we evaluate their competence.

2.2 Decision Making and Cognitive Bias

The *halo effect* was previously established as a factor that can potentially impede a candidate's fair unbiased assessment but is far from being the only one. Numerous other cognitive biases affect human cognition effectively leading to the decision-making process being severely skewed in certain areas. Cognitive biases are essentially a product of how humans construct their own versions of reality from the inputs they receive utilizing senses prone to perceptual distortion and inaccurate judgment. These biases have been inscribed in our brains through evolutionary processes and are believed to be features rather than shortcomings of our cognitive system [36].

The existence of cognitive biases can be explained with the concept of bounded rationality [37] which addresses the specific constraints agents are confronted with in their environments. The rational decision-making process is limited by certain inevitable constraints. Humans only have limited time, information, or even cognitive capacity to come to a decision. So, when picking out a mate, choosing what food to eat or what product to buy, we rely on a set of heuristics. Cognitive systems are assumed to be fundamentally adapted to their environments, according to the principle of bounded rationality. A common metaphor used in this context is that of the mind as an adaptive toolbox. The cognitive tools we have at our disposal are limited and have been tuned to characteristics of natural environments through evolutionary processes. Consequently, a cognitive bias is to be understood as the tendency to solve a given problem using a particular cognitive tool. The decision on what cognitive tool or heuristic is being utilized is a product of the mind's assumptions about the environment it finds itself in [38].

While cognitive biases may have once been critical for swift decision-making to ensure our survival in the wild, they can now lead to irrationality and discrimination. The following section details a selection of cognitive biases relevant to the hiring process and the workplace in general.

2.2.1 Affinity Bias

Affinity bias describes the tendency to favor people who share similar interests, have a similar social background, or share our own experiences. In the workplace, this can result in a very uniform workforce and is even justified and institutionalized by the term 'culture fit' which often translates to someone who shares the same experiences and the same background as the hiring team. Factors can range from playing the same sports growing up to attending the same university. Affinity bias also leads to stifled career opportunities for minority groups, because the men in leadership positions tend to favor candidates with similar experiences and consequently favor men when allocating career-enhancing opportunities [39].

2.2.2 Anchoring Effect

Whenever someone tries to estimate a numeric value, they use an anchor point as a reference. While this can help to come up with an educated guess in cases where the anchor is consciously and adequately picked, the anchoring effect works unconsciously and can severely throw off an estimation [40]. In 1974 Tversky & Kahneman conducted an experiment that had participants estimate the number of African countries that were members of the United Nations, but only after they had spun a wheel of fortune. Participants who landed a higher number on the fortune wheel tended to estimate the number of African UN member states higher than those who had spun a lower number [41]. This goes to show that we do not only use anchors that are related to the domain of the question at hand to inform our estimations. We are just as easily influenced by completely unrelated information. The anchoring effect is very robust and does not respond to many moderating variables. Even extreme values that can not plausibly apply to the estimation problem at hand still yield an effect [42]. The phenomenon is so robust that even when explicitly instructed to correct for the potential influence of an anchor, participants still struggled to mitigate the effect [43]. Compensation negotiations are only one of the aspects in the hiring process prone to be influenced by the anchoring effect – anchoring generall10 impacts decision-making [44].

2.2.3 Attribution Biases

Attribution describes a process by which we try to make meaning of the actions of others. We constantly make assumptions about why people might behave in certain ways and judge their behavior based on our attributions. These attributions, however, do not necessarily reflect reality. Perceptual biases play a major role in how we interpret socially relevant behavior leading to inaccurate assessment. One of the most important of these

biases is the fundamental attribution error. It comes into play whenever we try to make sense of someone else's behavior [45]. When people observe the behavior of others, they show a tendency to overemphasize the role of dispositional factors while minimizing the influence of situational factors at the same time. This holds true even when they are aware of compelling external reasons contradicting this hypothesis, as was observed by Jones & Harris [46]. In their 1967 study, they had participants listen to political speeches. The speakers had not written the speeches nor did they necessarily conform to the content of the speeches they were giving. Although the participants were fully aware of the fact that the speakers were merely reciting a speech they had been randomly assigned, they still attributed the attitudes that were transported in the speech to the speaker. But although people tend to overvalue dispositional factors when interpreting the actions of others, they overemphasize the role of situational factors when judging their own behavior [47]. Additionally, people tend to overestimate the role of their own contributions in joint products and will attribute their failures to situational rather than dispositional factors [42]. This can be detrimental in a job application context: An applicant arriving late to a job interview can easily be interpreted as being dispositionally unpunctual or not valuing the interviewer's time when the delay might in fact be caused by legitimate external factors outside of the applicant's control such as a technical defect.

2.2.4 Beauty Bias

The beauty bias or physical attractiveness stereotype is a tendency to ascribe competency or other socially desirable personality traits to physically attractive people. In their 1972 work 'What is Beautiful is Good' Dion et al. demonstrated the influence of stereotypes regarding physical attractiveness on the assessment of a person's character. Participants generally evaluated attractive people to be more honest and intelligent [48]. These tendencies translate into real-world scenarios where they manifest, for instance, in the form of a *beauty premium*. Hamermesh & Biddle found that more attractive people generally had higher salaries and that people with below-average attractiveness also received below-average wages [32].

2.2.5 Confirmation Bias

When searching for and interpreting information, confirmation bias negatively affects the chances of an existing hypothesis being rejected. We process information in a way that unintentionally fosters the immunity of the hypothesis by systematically seeking out or remembering only information that supports our hypothesis while disregarding contradicting information. Confirmation bias might even lead to systematic re-interpretation of information to make it compatible with an existing hypothesis [42]. The way our cognitive system works leads to an automatic bias, based on how we are presented with certain information. Kahneman demonstrates this with the following example. When asked the question, "Is Sam friendly?" our mind will typically conjure memories of instances of Sam being nice. When asked, "Is Sam unfriendly?", however, different instances of Sam's behavior will come to mind. Our mind will start a deliberate search for confirming

evidence to test the hypothesis. This so-called *positive test strategy* functions contrary to how hypotheses are tested in science by trying to refute them [49]. All this can lead to confirmation bias having a detrimental effect on how job applicants are evaluated. Once an interviewer has formed an opinion about an applicant, even though it is exclusively based on their name or their application photo, confirmation bias will likely have them see their opinion confirmed.

2.2.6 Contrast Effect

When presented with two contrasting pieces of information, the contrast itself leads to a distortion in judgment. When first putting your left hand in a bucket of cold and your right hand in a bucket of hot water and then putting both in a bucket of lukewarm water at the same time, the water will feel hot to the hand that had been in the cold water and cold to the other hand [40]. As humans, we are not equipped to make absolute judgments. Rather, we can only make judgments based on the current context. This holds true especially in the professional realm when considering job interviews where it was shown to have a significant impact, especially if applicants were seen in couplets with diminishing effects when seeing applicants in groups of more than four [50].

2.2.7 Halo Effect

The halo effect can influence a person's evaluation of an individual's performance or competence. However, in contrast to the leniency/severity error, it is not a constant error that applies to all subjects being evaluated. Rather, a single positive or negative aspect of a person can influence all other appraisal dimensions. This may result in individuals being evaluated to be competent in leadership, management, personnel administration, etc. just because they proved to be very cooperative [51]. The halo effect increases the weight of first impressions since whatever we first learn about someone has an impact on how we process subsequent information about them [49]. This fact is especially relevant in the hiring process where the person going through applications might judge a candidate's potential while being subconsciously influenced by the layout of their CVs or their application photo.

2.2.8 Leniency/Severity Error

The leniency and the severity error are two cognitive biases that can appear in rating scenarios. They describe a systematic error that occurs due to the tendency of a rater to evaluate participants' performance or abilities either consistently overly negative (severity error) or exceedingly positive (leniency error) [51]. The rater's bias consequently leads to undeservedly low or high and therefore inaccurate scores. In the context of performance evaluation in the workplace, this may lead to severely distorted employee performance ratings. While one manager might consistently be more lenient when evaluating their subordinates' performance, another manager might be overly strict resulting in performance ratings hardly reflecting an individual's actual performance.

2.2.9 Mere Exposure Effect

Repeated unreinforced exposure to something influences the attitude we have towards the said thing. Mere repeated exposure biases us towards a stimulus so that we develop positive attitudes based on sheer familiarity [42]. Most areas of human decision-making are affected by the mere-exposure effect. Investors tend to favor domestic companies because they are more familiar to them [52]. Academics evaluate journals they have previously published or reviewed for systematically higher than their peers without prior histories with the publications [53]. Even voting patterns were found to be correlated with a candidate's exposure irrespective of the popularity of their policies [42].

This has numerous implications for hiring and for the workplace in general. Organizations might stick to outdated strategies or sub-optimal processes simply because managers are familiar with them. The same principle applies to the adoption of new technologies [54]. The mere-exposure effect also helps to create and reinforce social norms and stereotypes [55] perpetuated by mass media.
On the other hand, however, mere-exposure can help to reduce biases against minorities. Zebrowitz et al. found that exposure to faces of other-race individuals can increase the liking for strangers of that race [56].

2.3 In Conclusion

This chapter has explored the nature of implicit prejudice and its impact on the workplace based on a paradigm shift in social studies leading to a better understanding of the nature of prejudice. We now recognize that implicit biases, unconscious associations formed through our experiences and environments, operate unintentionally and can negatively impact our decisions, particularly in professional settings. I have explained how unconscious biases can influence the hiring process leading to discrimination against qualified candidates from minority groups. Furthermore, I have discussed the role of cognitive biases inherent in human cognition in distorting perceptions and impacting decision-making in the workplace.

By acknowledging the existence of implicit bias and cognitive biases, we can begin to mitigate their effects. The following chapter will explore strategies and interventions that organizations and individuals can employ to promote diversity, equity, and inclusion in the workplace.

CHAPTER 3

Promoting Workplace Diversity

An ever-growing body of research suggests the benefits of diversity in the workplace on both company performance as well as workplace atmosphere. Workplace diversity can lead to better decision-making, greater creativity and innovation, and more successful marketing to different types of customers, increasing marketing opportunities and business image, enhancing organizations' abilities to compete in global markets [57, 58, 59]. A 2009 study examining a national sample of US for-profit business organizations found that diversity was associated with increased sales revenue, more customers, greater market share, and greater relative profits [9].

Although an increasing number of employers acknowledge the benefits of adopting a pro-diversity stance, the strategies for increasing workplace diversity can vary greatly. To address research questions RQ1 and RQ2, I will provide a brief overview of assorted methods commonly employed to promote workplace, and in particular hiring diversity in the following.

3.1 Targeting Diverse Applicants

One strategy employed by organizations to increase the diversity of their workforce is to specifically target minority groups by making a conscious effort to carefully word their job postings in a way that encourages diverse candidates to apply rather than deter them. Research has found that the usage of words associated with masculinity, also known as agentic language in job postings can hurt the number of female applicants. Terms like *challenge*, *lead*, *boast*, and *active* in advertisements made jobs appear less gender diverse and women felt they would not fit well in such a position or even company [60].

Especially in male-dominated industries, many cases of masculine-themed phrasing in job postings lead to women refraining from even applying for a job. A study examining job postings for internship positions in the financial sector found that women applicants

perceived a higher level of fit with positions and were more likely to apply to postings high in communal language and low in agentic language. Furthermore, the study showed that even employers included in top diversity rankings had their job postings phrased in agentic language and hence they contained gender bias [61]. In addition to paying careful attention to the wording of job postings, it can be beneficial to specifically target recruitment activities to underrepresented populations for instance by working with minority organizations, as is reflected, for instance, in a set of guidelines published for the University Health Services of Berkley, California [62].

3.2 Unconscious Bias Training

Training and workshops are popular ways of trying to address hiring discrimination and are at the heart of many diversity strategies. Most of these training efforts are designed to raise awareness for unconscious and cognitive biases and the role they can play in general and especially in a workplace environment. A whole industry of diversity consulting has emerged and diversity training has been widely adopted as a method. According to Harvard Business Review, virtually all Fortune 500 companies offer diversity training to their employees, usually without measuring its impact [63]. While there certainly is value in raising awareness for unconscious biases, research shows that the effects of diversity training are often negligible, especially for awareness training. Diversity training seems to be less effective in changing attitudes due to the attitudes that are supposed to be addressed being generally strong, emotion-laden, and tightly tied to trainees' self-identity, as was found in a 2016 meta-analysis of over 40 years worth of research on diversity training [10]. The same study found no compelling evidence that the long-term effects of diversity training are sustainable in relation to attitudinal/affective outcomes. In contrast, training focused on cognitive learning yielded effects that remained stable or even increased in the long term. The authors also found evidence for training effectiveness being influenced by additional initiatives complementing the diversity training, thus suggesting that diversity training is less effective as a singular measure. Some scholars even point out that there is no evidence base of effectiveness for unconscious bias training, stressing that there is no proven link between knowing about bias and changing behavior [64]. There is even evidence suggesting that mandatory diversity training can have the opposite effect, leading to less diversity rather than more – a phenomenon explained by the tendency to respond to compulsory courses with anger and resistance [65].

3.3 Blind Hiring

Blind hiring typically refers to the utilization of an anonymous job application process, most commonly limited to the initial job applications. Anonymous job applications do not include information about an applicant's identity and status as a member of a minority or disadvantaged group. What characteristics are left out or removed from a job application can vary depending on the setting. Typically they would include the applicant's name, picture, contact details, gender, and nationality as well as information

referring to their relationship status and their personal life. Effective implementation of anonymous job applications can level the playing field in access to jobs by emphasizing candidates' skills and qualifications. The approach exhibits some apparent benefits. Numerous correspondence studies provided evidence for the discrimination of minority and disadvantaged groups when it comes to receiving callbacks. By removing all information that has the potential to activate biases from the applications, it is possible to prevent discrimination at the initial hiring stage. Furthermore, recruitment decisions that are solely based on skills and qualifications should automatically lead to outcomes in line with a firm's objective to hire the most productive workers [66].

A popular example of the potential effectiveness of blind hiring are anonymous auditions for hiring musicians in symphonic orchestras. The research found that the likelihood of a female applicant advancing to the final round of an audition was increased by severalfold by introducing a screen separating the player from the committee and thus hiding the applicant's identity [11]. For the Boston Symphony Orchestra, the first Orchestra to conduct auditions behind a screen, the adoption of said procedure led to women's likelihood to advance rising by 50% [67].

While blind hiring is a promising concept, there are certain limitations to its positive effects. Introducing anonymous job applications can be associated with significant costs and can be a labor-intensive and error-prone task, depending on the methods used to de-identify applicant data. Furthermore, even with anonymous applications in place, the applicants' identities are going to be revealed at some point, so that improvements in job offer rates for minorities might be limited, even though they would profit from higher callback rates [66]. In some cases, anonymous applications can even decrease opportunities for minority candidates, as demonstrated in a recent study by the *Behavioural Economics Team of the Australian Government* [68]. Researchers found that *Australia Public Service* representatives positively discriminated against women and minority candidates by default leading to a decreased likelihood of callbacks for these candidates when applications were de-identified This speaks to the fact that anonymous job applications have the potential to reduce discrimination only when discrimination is high in the first place [66].

Blind hiring is anything but a common practice in Europe where application documents typically include much more information than is the case in, for instance, the United States. Many employers are still averse to the concept and the *Austrian Economic Chambers* even published a position paper focusing entirely on the downsides of anonymous applications [13]. Nevertheless, the debate about anonymous job applications shows an interesting approach European countries take toward the policy. Numerous field experiments have demonstrated the potential of adopting a policy mandating anonymous job applications. Findings, however, have yet to lead to any legal adaptions towards making anonymous job applications with all necessary preconditions the norm [69].

3.3.1 Text-based Communication

While text-based communication offers the greatest opportunities for anonymization, particularly in settings like online job applications or virtual interviews, it can also present unique challenges for self-presentation. Understanding these challenges is crucial for designing effective communication strategies in computer-mediated interactions. The absence of nonverbal cues in text-based communication can present itself as a major hurdle. Generally speaking, communication channels with a limited ability to convey nonverbal cues lead to a slower rate of exchanged information [70]. This means that a textual exchange might take more time to convey the same amount of information. Facial expressions, body language, and tone of voice play a significant role in conveying emotions, intent, and personality during face-to-face interactions [71]. Without these cues, the recipient of a text-based message may struggle to accurately interpret the sender's meaning. Nuances in sarcasm, humor, or emphasis can be lost in translation, potentially leading to misunderstandings [72].

Furthermore, text-based communication offers limited ways to express emotions compared to face-to-face interaction. Research suggests that emojis can serve some of the same functions as actual nonverbal behavior [73], therefore not only complementing but also enhancing verbal messages. However, emoticons and other forms of more informal language may not always be appropriate or accurately convey the intended emotion [74, 75].

The difficulty of conveying personality in text is another challenge. Text-based communication can make it challenging to showcase one's personality effectively leading to the recipient struggling to form a complete picture of the sender based solely on their written words [76]. This can be particularly disadvantageous in situations like online job interviews where a strong first impression is crucial.

The challenges that come with the absence of nonverbal cues are, however, accompanied by potential benefits: Without the need to monitor nonverbal expression, people can allocate more resources to the production of their verbal messages [77]. According to the hyperpersonal perspective, proposed by Walther [78], conversational actors are able to attain certain interpersonal goals through adaptation to the lack of nonverbal cues in CMC. Goal-enhancing messages are facilitated by the communication channel, due to the greater control over message construction than is available in face-to-face settings [77].

To conclude, text-based communication comes with its own challenges that need to be considered when implementing it in an interview context but also offers some distinct advantages such as the redundancy of monitoring nonverbal cues which can in turn free resources that can be redirected to verbal message production.

3.4 Algorithmic Hiring

A different approach known as *algorithmic hiring* promises to address fairness in hiring decisions while simultaneously minimizing employer effort. The increasing pervasiveness

of information technology has led to rising application numbers received by possible employers. This results in a significant amount of overhead when trying to fill an open position [79]. Making use of algorithmic solutions or AI systems can be a way of addressing the excessive efforts resulting from this fact. AI can be employed in nearly all stages of the recruitment process, be it to target a specific subset of individuals as potential candidates, to screen job applications, or to assess a candidate's viability regarding a specific position [80]. While these algorithmic solutions often promise to allow for a fair assessment unfazed by biases inherent in humans, in actuality, the matter is far more complex. Discrimination might start even before a person even applies for a job, for example by using targeted advertising tools on social media to reach out to possible candidates [81]. The implementation of algorithmic hiring comes with many pitfalls and does not necessarily lead to a reduction in biases but can actually fortify existing biases by systematically embedding them in the selection process, as witnessed in the case of a hiring tool developed by Amazon [82]. Algorithmic recruitment is relatively new yet consistently gaining interest. Organizations have to take implications such as ethics, labor law, and data privacy into account and have to plan for frequent checks to ensure that the AI system stays within acceptable operational boundaries [83]. Algorithms are nothing but human creations and might have inscribed in them the biases of the people who wrote them in the first place. This applies to AI as well through which algorithmic discrimination can be unwittingly propagated if the input data used to train the systems is not carefully selected. Additionally, it is remarkably hard for researchers to detect algorithmic bias in commercial algorithms since there are no incentives, much less requirements for the private sector to disclose how these algorithms are developed [84]. The use of algorithms can help to cement existing discrimination in particular. An analysis of an algorithm used by the *Public Employment Service Austria (AMS)* to assess the employability of their job-seeking customers found that belonging to a group that is structurally discriminated leads to a reduction in their predicted re-integration chances. This can potentially result in customers with lower predicted chances of reintegration receiving less support, and being refused access to training measures [85]. Optimising AI systems for fairness is anything but trivial. There are around 21 differing definitions of fairness in computer science alone, some of which are mutually exclusive. For example, providing equal opportunity without bias would be challenging to implement in an AI combining it with measures correcting for historical or inherent disadvantages or social injustices [83].

3.5 Concrete Examples

When it comes to state-of-the-art solutions implementing the aforementioned methods for addressing discrimination and fostering workplace diversity, there are a number of commercially available tools, each focusing on different aspects of the hiring process. The current section provides an overview of (commercially) available software products that claim to offer solutions to the bias problem that can occur with job applications. While they each follow different approaches to address biases in distinct stages of the application

Methods	Products
Anonymized applications	Pinpoint, Blendoor, Entelo, TalVista
Skills/Personality tests	Toggl Hire, Pymetrics, Interviewing.io, Applied
Anonymous interviews	Interviewing.io
Job postings	Gender Decoder, Textio, Datapeople, Applied, Ongig

Table 3.1: Overview of available tools that offer features to mitigate bias in hiring, grouped by method of mitigation

process, a comparable solution addressing the focus of this research, namely a typical job interview, does not appear to be available at the time of this writing. Table 3.1 provides an overview of the examined products.

3.5.1 Job Postings

As laid out in Section 3.1, the wording of job postings can be a source of discrimination. Certain terms commonly used in job postings can play into the unconscious biases of the people who are in the job market and consequently impact who applies for the job in the end [60]. Specific coded words have been identified that encourage some applicants to apply while simultaneously discouraging others. This phenomenon has been recognized by a host of organizations. Google, for instance, has reported an 11% increase in applications from women in their 2020 annual diversity report [86], after deciding to address this fact by introducing an internal software tool that removes words and phrasing that could bias a candidate against applying for job advertisements.

While, strictly speaking, measures to word job advertisements in a way that does not discourage specific groups of people in the job market from applying may not be subsumed under the umbrella of *blind hiring*, they are in fact a prerequisite for the measures employed later in the hiring process to work. In the following, I will discuss several software tools intended to solve this problem by screening job advertisement copy for coded language.

Textio is a web-based service that operates on a subscription-based model, offering to analyze all kinds of brand communication in regards to, among others, gender-coded language [87]. Having been built by a team of experts for machine learning and linguistics, *Textio* was created with a technological, rather than a social agenda. It is not based on static research, but was first trained with a combination of information scraped from job sites and information about subsequent hires contributed by an early list of participating companies [88]. The software uses statistical information to determine if certain words used in a job posting or recruitment email are likely to speak more to men than women and vice versa and provides the user with alternatives to substitute coded phrasing [88, 89]. The overall tone of the text can be analyzed and displayed on a scale with a completely masculine tone on one and a feminine tone on the other end. Additionally, *Textio* also provides information on how likely a job posting is going to appeal to a specific age group. This way employers are able to tailor their communication

to their intended target demographic, which might differ depending on the professional field or, as *Textio* co-founder Jensen Harris puts it: "In tech, people want to hire more women, but in nursing, they want to hire more men." [88]. *Textio's* interface overlays its analysis by color-coding words and phrases in the text. Words that tend to encourage men more than women are highlighted blue, repetitious phrasing is marked grey, and technical jargon is coded red. The software also offers inclusion guidance concerning age and ability bias and promises to help avoid exclusionary metaphors and harmful language.

While *Textio* uses machine learning techniques to detect biased language, there are quite a few alternatives that use a more traditional approach to analyzing text. The freely available *Gender Decoder* is the most bare-bones out of the bunch. Based on the word list proposed by Gaucher et al. [60], the online tool is free software published under the MIT license [90] and detects gendered wording.

The language analytics for job posts offered by *Datapeople* is another web-based service that helps employers with their job descriptions and applicant communication [91]. Like *Textio*, the service highlights words and phrases that are likely to deter specific groups of potential candidates. Additionally, the software provides a bullet point list of suggestions to improve the advertisement based on its analysis. According to *Datapeople*, their tool is able to detect language that conveys sexism, racism and tokenism, ableism, ageism, elitism, and religious bias.

The job description tool available from *Applied* offers similar features for the analysis of job postings. Other than its competitors, though, it includes a feature that helps to streamline job requirements by removing low-priority items that could potentially discourage women in particular from applying [92].

The *Ongig* text analyzer software [93] is yet another tool that allows employers to have a bias in their job descriptions detected and provides them with inclusive alternatives. Users are provided with an overall gender neutrality score for the text

3.5.2 Job Applications

The products presented in the following are examples of blind hiring as laid out in Section 3.3 put into practice. While I found no solution providing a pure implementation of blind hiring where a candidate would remain anonymous up until the point where they are hired, all of the tools feature aspects falling under the blind hiring umbrella, such as redaction and/or pseudonymization of applicant data.

Pinpoint is described by its creators as an 'end-to-end talent acquisition software'. It includes features to attract candidates such as branded careers websites or programmatic recruitment advertising and allows for the creation of custom online application forms. Furthermore, the software promises to help with the automation of tedious recruitment processes, basically working as a platform solution that can model and administrate every step of the recruiting process, even boasting features to help with onboarding new hires.

Pinpoint also sports a feature called 'blind screening' that promises to eliminate bias by anonymizing applications and resumes and thus ensuring that 'hiring teams assess applicants based exclusively on their experience and skills'. The feature is focused on the early recruiting efforts leading up to the interview stage. The software parses applicants' resumes by default, creating standardized versions of applicants' CVs for every candidate including their personal information. The 'blind screening' feature can be enabled on a per-job basis and, once active, hides the applicants' personal data. Every candidate is assigned a randomly generated alias made up of a combination of colors, fruits, and vegetables. Data that might be available about a candidate, such as their contact details, their native language, or their age are hidden from the recruiting staff by the software when 'blind screening' is enabled. *Pinpoint* even claims to remove indicators for potential protected characteristics from free text answers given in the application form [94].

Blendoor [95] is a self-described social impact analytics company offering AI-supported services to companies and individuals. Their proprietary BlendScore technology uses a data-driven approach to analyze corporate DEI (diversity, equity, and inclusion) performance. The company started, however, with software that matched diverse job seekers with employment opportunities by capturing applicants' profiles from existing online job boards and applicant tracking systems and stripping personal data from them [89]. The system hid a candidate's name and photo as well as any indication of their age, only showing their skills, work experience, and education. The founder of the company additionally stated that the software had a layer of accountability built in. Tracking how far certain demographics of people made it in the recruiting pipeline, she suggested, would make decision-makers more conscious about what they were doing [96]. The original *Blendoor* app seems to have been discontinued as the service is no longer mentioned nor advertised on the *Blendoor* web presence.

Entelo [97] is a one-stop recruiting solution with an explicit focus on diversity. Not only does it offer the so-called 'Unbiased Sourcing Mode' that hides and anonymizes candidate information commonly associated with bias, but it also offers filters specifically intended to find members of underrepresented groups in the candidate pool. This feature is particularly helpful for organizations that seek to increase diversity in their workforce by explicitly hiring for diversity. The *Entelo* web presence does not go into great detail about their anonymization feature, but the example they provide on their website shows that photos and gender are removed from the candidate information, and applicant names are reduced to initials. Other information that might have an adverse impact, such as the name of the educational institutions, remains available.

TalVista's redacted resume feature can process applicant CVs and redact personal information such as names, photos, and birth dates from the documents while still preserving the original layout, which according to *TalVista*, allows candidates 'to choose how they present themselves, keeping format, presentation, and personality intact' [98].

3.5.3 Skills and Personality Tests

Toggl Hire is a platform offering their customers skills tests to screen candidates, promising to cut the time to hire by over 80%. *Toggl Hire* claims to provide a rapidly growing library of peer-reviewed tests that are taken by thousands of candidates every week. It is possible to use a provided test from the library as is or to amend it to taste. Creating custom tests is another feature offered by the software. The skill tests are taken remotely via a web interface and are purposefully time-pressured. Candidates can be filtered by their test scores, a feature intended to help with shortlisting. A collaborative assessment of candidate scores is made possible by allowing to leave notes, tags, and ratings on candidates for other team members. *Toggl Hire* claims to boost hiring diversity by creating a merit-based system, liking hiring to sports where 'the best person should win - regardless of their age, where they come from, or where they went to school'. The usage of skills tests over resume-based evaluation is intended to show candidates that companies using the software value 'abilities over fancy colleges or impressive corporate histories' [99].

Pymetrics is an example of using machine learning in conjunction with scientifically valid methods of assessment to enhance the recruitment process. The company uses neuroscience assessments to match applicants with specific jobs, providing an unbiased hiring process. A gamified brain test routine is used as the first point of contact for applicants, effectively replacing the resume as a first pass-filter [100, 101]. The results candidates receive when taking the *Pymetrics* tests are benchmarked against a custom profile that was automatically created by having the companies' top performers in comparable positions take the same tests. This results in a customized algorithm used to evaluate the candidate's performance. This can inherently be a source of bias itself due to historically inequitable hiring practices, the algorithms are monitored and adjusted by *Pymetrics* for each company. They know the gender and ethnicity of the reference group and can adjust the model to correct disproportionate results in cases where, for instance, men receive uniformly higher scores than women on a given trait [101]. These features are consequently removed or down-weight if they are found to be highly correlated with the protected attribute in question [102]. *Pymetrics* claim that their system is proven to be free of gender and ethnic bias [103] and seeks to ensure this by pretesting their algorithms [104] and establishes additional accountability by open-sourcing the tests they use [102]. One significant limitation to the methodology used by *Pymetrics* is that it only works for sufficiently large organizations. To generate reliable results, the models need to be trained by ca. 100-150 current employees playing the games. For organizations that have successfully integrated *Pymetrics* into their hiring process, the company reports significant efficiency gains in the form of a 90% reduction in time to screen applications and a 62% increase in female representation for talent acquisition, as was observed in case studies [105].

Interviewing.io is a platform that allows candidates to practice technical interviews anonymously and online with former tech executives [106]. The company collects audio transcripts, data, and metadata describing the code the interviewee writes in the course

of the interview along with detailed feedback by the interviewer as well as the interviewee [89]. Once a candidate has gotten enough practice, they can be invited to interview anonymously at tech companies, directly skipping to the tech phase of the interview process. This way *Interviewing.io* allows candidates to skip the initial in-person screening – a potential source of bias in the traditional application pipeline. If a candidate feels they have done well in an interview, they are free to reveal their identity to the interviewing organization and proceed to the next stage of the recruitment process, which typically is an on-site interview. Thus, potential bias is not completely removed from the hiring process, but candidates are given the chance to leave an unbiased first impression before they reveal their identity [106]. While the process used by *Interviewing.io* eliminates common biases, the validity of the technical interview as an assessment tool is called into question by a study examining over a thousand interviews using *Interviewing.io* that found performance to be arbitrary from one technical interview to the next. The study concluded, that interview performance is volatile and only a minority of candidates are consistent in their performance [107, 108].

Complementary to their job description tool, *Applied* [109] also offers features to anonymize applicant information. *Applied* moves the focus away from the resume by assessing candidates based on their responses to work-based scenarios. The candidates' answers are then stripped from all personal details and chunked up to facilitate comparative assessment, removing all but the most predictive, thus relevant, information. The answers are then randomized to be scored and averaged across multiple independent assessors [67, p. 35].

3.6 In Conclusion

This chapter explored the impact of a diverse workforce on a company's success, outlining a framework for achieving diversity through targeted recruitment, unconscious bias mitigation, and leveraging technology. The presented techniques each pertain to different aspects of the job application process and come with upsides and potential shortcomings. Finally, these measures were put into perspective by providing an overview of available tools employing the presented techniques for promoting diversity in hiring. There is a host of solutions available to companies trying to fill an open position while trying to avoid discrimination. However, none of the available tools addresses the common practice of job interviews presenting the focus of this research.

Part 2

Design

CHAPTER 4

Pre-study

To be able to design a system that would increase fairness in hiring processes that would be accepted by the stakeholders in the hiring organizations, it was necessary to gain an impression of what some of the typical hiring and diversity practices looked like in Austrian companies. For this reason, a pre-study was conducted in which people involved in recruiting for six different companies were interviewed. The span of organizations ranged from a medium-sized software business with under 200 employees to a large internationally operating financial institution employing more than 260 times that amount. This would serve as the foundation of the design process and consequently, the findings detailed in Section 4.2 would help to answer research question RQ1.

The interviews were qualitatively analyzed to extract information not only on the current recruitment processes but also on the role diversity plays in organizations and what measures are being taken to promote it. Company representatives provided their views concerning unconscious bias in hiring and anonymization as a potential mitigating measure. Since it was clear that the system that was to be designed would incorporate at least some level of anonymization, possible past experiences with and general opinion on anonymization and de-identification technology were a particular focus of the interviews.

4.1 Methodology

The methodology employed to conduct the pre-study will be laid out in the following including the research context, method of data collection, and the approach for data analysis. The pre-study would include a round of expert interviews in a group setting with representatives of all companies scheduled to partake in evaluating the first iteration of the DEBIAS prototype. The recorded interviews were transcribed and preliminarily analyzed by means of Mayring's qualitative content analysis [14].

4.1.1 Research Context

The research for this thesis was conducted in the context of a set of recruiting events organized by the *TU Career Center*, an organization affiliated with the *Vienna University of Technology* which seeks to facilitate the entrance into the job market for students. To achieve this, they offer a set of services for students, helping them with their application documents, preparing them for job interviews, or connecting them with businesses that have open job postings. The *TU Career Center* has been hosting different event series, allowing students low-threshold access to representatives of businesses looking to hire and vice-versa, for years. When recapitulating their experiences with these kinds of events, they discovered that students with a migration background seemed to be much less likely to be hired. Based on this impression, the *TU Career Center* decided to develop a new concept for an event, designed to take this apparent discrimination into account by removing all factors that might potentially activate bias. After drafting a preliminary concept, which stipulated separating students from company representatives via a screen and establishing anonymity by distorting the applicants' voices, they decided to get the *Centre for Informatics and Society (C!S)* on board to provide scientific guidance for the project. The *C!S*, and hence the author, would develop the concept for the event along with the technology needed and collect scientific data during the event, while the *TU Career Center* was tasked with the organization of the event, as well as recruiting students and companies to participate.

4.1.2 Expert Interviews

A set of qualitative semi-structured expert interviews was conducted with representatives of all six organizations participating in the event. An overview of the surveyed companies including their respective trades, their approximate number of employees, as well as who represented them in the interviews can be found in Table 4.1. The examination was carried out in the form of group interviews taking place in company facilities, each featuring two to six interviewees.

One interview session per company was scheduled, lasting about 34 to 69 minutes with an average interview length of just over 53 minutes. The interviews were recorded to be transcribed by a professional academic transcription service for later analysis.

Interviews, as compared to other methods of gathering verbal data, are best suited to collect data that helps understand people's perceptions and experiences [110]. The information gathered in the interviews was intended to provide insights into the recruitment practices currently employed by the participating businesses, representing a sample of Austrian businesses in general. The design of the system and process would take the current practices into account, allowing for theoretical integration into existing processes. Another important factor was to ensure acceptance among the personnel involved in hiring. Gathering opinion data on what recruiters and other decision-makers considered to be crucial information to have to assess a candidate made it possible to design a

process that would accommodate these factors and therefore ensure a certain level of acceptance.

The semi-structured approach was selected to allow for possible unanticipated expansions of the topics that were discussed. Since there was no clear-cut image of how the organizations handle their recruiting, the semi-structured interview allowed for additional exploration. An interview guide was created, including not only questions addressing possible anti-discrimination measures in the hiring process but a range of contextual information as well. Company representatives were asked to describe the typical hiring processes employed by their organizations, including the involvement of technology in each process stage. Additionally, they were asked to provide insights about their experiences with the theme of *diversity* and the measures being taken by their organization to address possible discrimination. Finally, interviewees were asked to design a hiring process themselves, that would account for biases and minimize or eliminate their role in the hiring process to prevent discrimination. The interview guide can be found in Appendix B in its entirety.

4.1.3 Qualitative Analysis of Interview Data

The expert interviews were transcribed and subsequently analyzed following a qualitative content analysis approach as proposed by Mayring [14].

Qualitative content analysis is a research method that involves the systematic and interpretive analysis of texts, images, or other forms of communication. It is commonly used in the social sciences and other fields to understand the meanings, attitudes, and perspectives expressed in a variety of sources.

According to Mayring, one of the key principles of qualitative content analysis is that it is an inductive and interpretive process that involves the identification and categorization of patterns and themes in the data. Mayring suggests a four-step process for conducting qualitative content analysis [111]:

Familiarization with the Data This involves reading and re-reading the data multiple times to gain an overall understanding of its content and context.

Coding This involves identifying and labeling key concepts and themes in the data, and creating a coding scheme to categorize and organize the data.

Reduction and Interpretation This involves reducing the data to its most essential components, and interpreting the meaning and significance of the patterns and themes identified in the data.

Verification This involves checking the validity and reliability of the analysis by comparing the results with other sources and through peer review.

Mayring stipulates the definition of three entities as the first step in the analysis to provide an underlying structure to the coding process. The *coding unit*[1] represents the smallest part of the text that can be coded, the *context unit*[2] defines the largest component of the text that can be assigned a single code, and the *evaluation unit*[3] defines which parts of the text will be analyzed and in what order. For the purpose of this study, the units have been defined as follows:

Coding Unit Any segment of the text as small as a single word shall be evaluated.

Context Unit A code shall not be assigned to any textual component larger than a set of associated sentences uttered in the context of an answer to a single interview question.

Evaluation Unit All parts of the text in answer to an interview question and that are on-topic shall be evaluated in chronological order.

Inductive category formation was employed to efficiently analyze the body of collected data. The method aims to arrive at summarizing categories that come from the material itself as opposed to predefined categories that were derived from existing literature. In contrast to the summarizing approach, not all of the collected material is taken into account for analysis. Rather, only those parts relevant to a specific research question are considered. Inductive category formation involves the definition of a theme or selection criterion for the selection process as a deductive element. Mayring describes eight distinct steps of inductive category development, which is a process with some iterative characteristics. Once a selection criterion has been established, the material is examined line by line, determining if it fits the criterion. If it does, it needs to be categorized, checking if it can be subsumed under an existing category or if a new category has to be formulated. Once a good portion of the material (Mayring suggests 10-50%) [112] has been processed in this manner, no new categories are to be found and the category system should be revised, ensuring that category boundaries are clear and the level of abstraction is adequate to the subject matter and the aim of the analysis. Mayring suggests generating 10 to 30 categories as a rule of thumb, stating that it can be beneficial to bring categories into order by formulating main categories. Once this categorization process has been completed, the resulting category data can be interpreted using two different approaches. Either the whole system of categories can be interpreted in terms of the aims of the analysis and used theories, or links between categories and passages in the material can be analyzed quantitatively, effectively examining which categories occur most frequently in the material [112].

Since the main focus of the analysis was on recruiting processes and how they might affect the chances of diverse applicants, the themes *recruiting* and *diversity* were established as primary selection criteria for the material. All in all, the inductive category formation

[1] Originally: “Kodiereinheit”
[2] Originally: “Kontexteinheit”
[3] Originally: “Auswertungseinheit”

#	Trade	Employees	Representatives interviewed
C1	Software development	180	Managing Director, Assistant to the Managing Director
C2	Financial services	47.284	Head of Recruiting, Diversity Officer, Diversity Associate
C3	Logistics	20.000	2 Recruiters, Diversity Officer
C4	Pharmaceutics	4.500	Head of HR, HR Associate
C5	Technology development	2.200	Head of Recruiting
C6	Infrastructure	2.185	Diversity Officer, Recruiter

Table 4.1: Overview of the companies surveyed in the pre-study

process initially resulted in 48 categories which were then consolidated by forming more general main categories subsuming the more specific categories.

4.2 Findings

All interviewees showed openness and general sensitivity toward the topic of diversity. Consequently, the value and benefits of a diverse workforce were acknowledged by representatives of all organizations. Still, representatives of almost all participating organizations took a rather defensive stance when asked about possible biases in their hiring processes. They would often acknowledge the role of personal information as a potential source for bias all the while reassuring that these biases did not play a role in their company. While diversity training was the most commonly mentioned measure of addressing potential discrimination, and the awareness of the existence of unconscious bias was generally in place, the effects were often estimated to be negligible by interviewees. In the following, I will provide a more detailed roundup of the most relevant topics discussed in the interviews.

4.2.1 Diversity

The surveyed organizations demonstrated nuanced interpretations of the topic of diversity. Out of the 6 organizations, 3 emphasized diversity being a subject needing and warranting an active commitment. While 4 companies claimed to follow a strict top-down approach in promoting diversity, one company, in addition to addressing it from the top-down, tried to promote diversity from the bottom-up in the form of self-organized initiatives by employees. The company had a set of special interest groups, for instance, to raise awareness of neurodiversity or LGBTQ issues that received company funding and resources. Representatives of two additional companies made mention of incentive structures that were implemented by their organizations to further gender balance. Gender balance, in these cases, was one of the key figures measured to assess management performance and had an impact on bonuses received by management.

One company displayed an interpretation of diversity focused on cultural aspects, mentioning that diversity was baked into the company culture because of the history of the company being intercultural right from the day it was founded. The same company expressed that there was no particular focus on gender issues when it comes to their diversity efforts. The remaining five companies all mentioned gender diversity as an explicit goal of their diversity strategies. However, when it comes to ensuring diversity, diversity training remained the only explicit measure to address potential biases, especially in the hiring process, in any of the companies. All but one organization offers diversity training as a part of their executive training programs.

4.2.2 Recruitment Processes

When filling open positions and recruiting new hires, all companies relied on job postings. In addition to publishing job offerings on their company web presence, all organizations specified to post job openings on jobs portals such as *karriere.at* or *jobs.derstandard.at*. Additionally, one of the companies mentioned a cooperation with a consulting firm specializing in disability management they had in place to specifically target candidates with disabilities. All of the surveyed companies do receive job applications via a dedicated online tool, while only a single company additionally accepts applications via e-mail as an alternative channel. Representatives of all surveyed organizations expressed that a CV is expected from applicants when it comes to the documents they need to provide. Only one company requires a letter of motivation from all of their applicants, however, representatives of two additional companies mentioned that they still liked to have them in special cases, for instance, to understand the motivation behind a candidate's decision to change jobs or if a CV does not perfectly line up with the job profile.

When asked if the hiring tools included means of anonymization all companies negated. All organizations require candidates to upload their CVs in document form when applying (minus one offering an option to source that information from LinkedIn as an alternative) and do not offer a dedicated interface allowing one to enter one's CV in a structured form. Representatives of one organization explained that they wanted to lower the effort required by applicants because they feared that applicants would not apply otherwise.

The first steps of the recruitment processes employed by all surveyed organizations are fairly similar. All companies generally have a two-tier interview process. Representatives of one company mentioned that the process could be adapted and tailored to the position to fill, meaning that special job openings such as management positions might include additional interview rounds or skills testing. However, skills testing was not used by any of the organizations as a basis for deciding who to invite for job interviews. This decision is entirely based on the documents provided by the applicants, such as CVs, motivational letters, and in some cases certificates for all organizations.

Once it has been decided on who to invite for a job interview, the interviews themselves are conducted in a semi-structured way. Structured interviews, which have been found to fare superior in promoting equity and facilitating a fair assessment for all candidates,

are not used by any of the surveyed organizations. Representatives of three companies mentioned that the natural flow of the conversation was taken into account when assessing candidates and a fixed structure would prevent a conversation from naturally unfolding. Furthermore, none of the surveyed companies use a fixed set of criteria specified in advance to assess applicants' performances. The job profile was mentioned by representatives of a single organization to function as a benchmark when evaluating job interviews.

When it comes to what information is of particular relevance in a job interview, other than sheer technical skills, representatives of 4 organizations mentioned that they tried to gather an impression of how motivated a candidate is. Additionally, all organizations claimed to assess how well a candidate will fit into a team, based on the interview. The evaluation of how well someone will fit into a team or a company is also known as *cultural fit*, a concept that can be vulnerable to bias. In extreme cases *cultural fit* is simply a moniker for social and cultural homogeneity, thus, stifling diversity by promoting like-mindedness [113].

4.2.3 Anonymization

None of the organizations reported using any form of anonymization in their recruitment processes. While applicants are not required to provide photographs of themselves by all but one organization, all six companies allowed the upload of documents including photos or had explicit fields for uploading an application photo in their online recruitment tools. Other data potentially able to activate biases, such as name, age, gender, and nationality is also available to staff administrating the pre-selection of candidates for interviews in all organizations. What data is required by applicants to complete an application form differs, however. While one company only requires the applicants to input their name and a means of contact, another additionally asks for their gender, their birthday, and their nationality. Yet, even when some personal information is not necessary for completing the application process, all of the described processes included the upload of a personal CV (or optionally the sharing of one's LinkedIn profile in one case), which, as was stated by multiple interviewees, usually include a lot of personal data including an application photo.

The idea of introducing anonymization in the first phase of the recruiting process was generally met with skepticism but also curiosity. One argument against it that was brought forth was that anonymization efforts would be in vain because even when an applicant's age was hidden from the recruiter, they could still infer that information from data extracted from their CV, such as education and employment periods. When asked what information they could imagine being removed from an application without it impairing their judgment, they mentioned photographs, gender, age, personal status, and children. Typically information that is not mandatory for the completion of the application but can be disclosed optionally was regarded as being negligible when it comes to bias and discrimination. There were multiple instances of company representatives acknowledging the potential biases that might be introduced by disclosing applicants' personal information only to later justify their current recruiting practices. One person

said, for instance, that they needed the photographs as a mnemonic device because otherwise, they would find it hard to recollect an interview and hence would have to rely on their notes alone. A representative of a different organization mentioned preferring to have an applicant's photograph and name available. They specified to judge applicants based on their address to see how far they lived from the company site all the while assuring that they were not influenced by said data due to their prior training and mindset.

Although much of the information applicants are asked to provide is considered optional, applicants themselves usually don't go to great lengths to reduce the amount of personal information they provide. A trend towards supplying unsolicited links to applicants' social media accounts or web presence in the application documents was noticed by recruiters. The consensus was that job applications including attached photos were the rule rather than the exception even when they were not required. One interviewee even spoke of self-discrimination, by which they referred to cases in which candidates provided additional information that could be interpreted to their disadvantage without the information being required to complete the application process.

CHAPTER 5

Design

The aim of this work is the development of a prototypical sociotechnical system to eliminate specific biases in the job application process with a special focus on the job interview. Since the system was expected to be used and evaluated in a specific real-world scenario, namely a recruiting event organized by the *TU Career Center*, key requirements of the design hailed from the cornerstones that had been laid out for the concept of said event. As a title for the project, the name DEBIAS, an acronym for *"Digitally Eliminating Bias in Applicant Selection"* was selected.

The *TU Career Center* reported that they had noticed patterns of discrimination in their traditional recruiting formats, claiming women and applicants with "foreign-sounding names" had far worse chances of being accepted by an organization, regardless of their qualifications and grades. To address this issue, they created a new format that would be specifically focused on eliminating biases in recruiting by hiding as much of the applicants' personal information from the company representatives, as possible. At the same time, company representatives should still be able to conduct job interviews that in some way resemble what they were used to, to garner acceptance among potential users. The current chapter addresses research questions RQ1 and RQ2 by laying out the core concepts and requirements that were identified for the process and software design and how they were established based on the insights gathered in the literature review and the pre-study.

5.1 Design Workshop

Following a user-centered design approach [114], the data gathered in the pre-study in combination with the requirements laid out in the event concept by the *TU Career Center* functioned as a jumping-off point for a design workshop held with HCI and CSCW scholars and UI experts in February 2020. The workshop, held at the *Multidisciplinary Design and User Research Group* of the *Institute for Visual Computing & Human-Centered*

Technologies was intended mainly as a brainstorming session to lay out the cornerstones of the system design, which would then be refined iteratively to produce the final functional prototype.

Outline Participants first received a short introduction detailing the project goals. They were told that a prototypical socio-technical solution was to be developed to pseudo- or anonymize applicants in a recruiting setting, that would take the needs of recruiters and companies into account and would be tested in the real-world setting of an event organized by the *TU Career Center*. After that, they were presented with an overview of relevant information on unconscious bias. An emphasis was put on the fact that unconscious bias operates without a person's awareness and is nothing that can be controlled at will so that even if recruiters are trained to be as objective as possible, their sub- and unconscious preferences will influence their decisions. The role physical attractiveness plays in hiring decisions and pay rates was discussed as well as the detrimental effects sequential evaluation of candidates can have on diversity. Furthermore, participants were made aware of the culturally normative nature of recruiting settings leading to looking, sounding, or behaving outside the contextual cultural norm being a detrimental factor for an applicant's potential success at being hired. Participants in the workshop were informed about the downsides of unstructured interviews, which, despite demonstrably being ineffective when it comes to evaluating applicants as well as combating bias, are still common practice. In the same vein, they were presented with an overview of alternative techniques recommended to minimize the potential impact of biases in hiring: removing personal information, structured interviews, and comparative evaluation of applicants. Finally, participants received a brief excerpt of the findings from the pre-study, informing them about the interview practices of the surveyed companies and their reservations concerning the anonymization of applicant data in the application process (cf. Section 4.2.3). The slides used for the introductory presentation given to the workshop participants are available in Appendix H.

Results In the course of the design workshop, it became clear that the design process could not be limited merely to the design of the system itself, but had to be expanded to the development of a process embedding the system in the larger context of the job interview and the event during which it was planned to be utilized. Ultimately the workshop yielded the concept of a two-tiered mode for the recruiting interviews as a compromise between a strictly structured and the more or less unstructured approach typically used by the organizations participating in the event in their conventional job interviews. The developed process integrates two separate modes of rating applicant performance, quantitatively for comparability on one hand, and qualitatively through the textual description on the other, and is intended to marry established recruiting procedures and practices with the technological support provided by the system.
The resulting concept outlined a process in which applicants would first be interviewed anonymously in a text-based format, where company representatives would go over a fixed set of questions with them and rate their performance on a per-question basis. Once

this has been completed, applicants would then meet the same company representatives for an open face-to-face interview.

5.2 Further Refinement

Once the fundamentals of the process yielded by the design workshop were laid out, the idea was pitched to the *TU Career Center* in a second workshop designed to identify the technical requirements the system had to meet in order to be usable in the context of the planned recruiting event.

The event itself was outlined to involve 6 companies that would each be interviewing at least 12 students, depending on how many students would sign up for the event. This meant that each interview would be strictly limited when it came to its duration. In conclusion, it was agreed upon that the whole interview process per company and student should take no longer than an hour to accommodate all interviews in the course of one event. The agreed-upon segmentation of each interview would consist of 30 minutes for the text-based structured interview, 15 minutes for the unstructured face-to-face portion, and 15 minutes of buffer set aside for the students to change location and potential unforeseen technical difficulties.

This led to a crucial decision in the design process. To allow for a meaningful exchange through the textual exchange in the short amount of time that would be available, it was necessary to provide a basis for the conversation. Since typing takes more time than speaking, the time slots set aside for the textual exchange were considered to be too short for the applicants to meaningfully answer more than one or two questions, which would be too little to give the company representatives enough information to gain a well rounded impression of an applicant. To address this, the process was adapted to include an additional step allowing applicants to textually answer a questionnaire provided to them by the company beforehand. The answers given by applicants would then provide the basis for the conversation. All applicants would answer the same questionnaire per company, functioning as the structure of the textual interview. Applicants' answers would sequentially be revealed to the interviewers only during the interview, meaning interviewers would not get to read the answers beforehand and hence be prevented from going into the interview with a preconceived opinion about the applicants.

This meant that the system had to provide additional functionality to the applicants, allowing them to answer the questionnaires, preferably at home. This fact was critical in informing the decision to implement the system in the form of a web application. This would not only allow students to submit the answers to the questionnaires in advance, but also meant that the tool didn't have to be explicitly distributed to the computers used in the event, but could simply be accessed via a standard web browser, a fact that would turn out to become crucial to the success of the event later in the development process.

To exhaust the possibilities of addressing bias in the format, the *TU Career Center* planned

to include workshops addressing unconscious bias and providing auxiliary information to participants, for students as well as company representatives. These workshops were intended to further raise awareness for biases on both sides and would prepare participants for the event, helping students to answer the questionnaires in a way that would not reveal information that could potentially activate biases. Companies, on the other hand, received guidance on how to construct their questionnaires and what follow-up questions to better avoid during the chat interview.

5.3 Core Concepts

The analysis of the pre-study findings and the results of the design workshop in combination with the concept that was outlined for the event resulted in a set of requirements for the system and the process it would be embedded in. The tool was designed to be a socio-technical hybrid – a mix of social interventions, namely workshops for participants on both ends, whether they were company representatives or applicants, and a technological infrastructure to support and structure the recruiting process. The technological and social aspects were designed to complement each other and work in conjunction to achieve a reduction of bias and promotion of diversity and fairness in the process. Three core principles were identified that represented the foundation of the subsequently developed design.

Structure Human resource managers often are firm believers in unstructured interviews. They generally rate them higher when it comes to perceived effectiveness than, for example, aptitude tests, personality tests, or general mental ability tests, all of which were found to be far superior to unstructured interviews when it comes to applicant evaluation [67]. Still, HR professionals rely on their intuition instead of probabilistic decision aids, a fact that was dubbed "the greatest failure of industrial and psychological psychology" by renowned psychologist Scott Highhouse [115]. Behavioral economist Iris Bohnet stresses that investment in structured interviews pays since they are a far better tool for the recruiting process [67].

In alignment with this finding, the interview process was designed to sequentially cover a predefined set of questions. Company representatives (henceforth referred to as recruiters) first define a questionnaire, which will then be answered by applicants in writing before the actual chat interview takes place. The answers given by the applicants beforehand will then serve as a basis for the actual job interview conducted on-site. Recruiters are not able to review an applicant's answers before the actual chat interview is conducted. The applicants' answers are revealed to the recruiters one by one in a guided process, according to an order previously defined by the company that remains the same for all applicants. After an answer given by an applicant has been assessed, the recruiter will ask the applicant follow-up questions based on their prepared answer. Once a recruiter concludes that a question has been sufficiently discussed, they rate the applicant's performance

and move on to the next question. This simple process is repeated for all applicants so that a set structure is maintained for all interviews.

Comparability Structured interviews only unfold their full potential, when they are paired with a comparative evaluation. Following the guidelines outlined by Iris Bohnet in 'What Works?' [67], the applicant performance is rated using a scoring system on a per-question basis to combat potential influences of the *halo effect* (cf. Section 2.2.7). Company representatives assign a weight value to each question signifying how important they deem the question in the context of the questionnaire. The assigned weight appropriately impacts the point value an applicant will be able to receive for an answer.

Only when the interview process has been completed for all applicants, recruiters can rank the applicants based on the impressions they have gained during the interviews. The ratings made in the course of the interviews represent the basis for comparing applicants' performances. This is intended to be a valuable tool to aid the decision-making process and aims at each of the candidates getting a fair chance at the job by providing access to information about applicant performance on a per-question basis, mitigating the influence of cognitive biases, such as *recency bias* and the *halo effect* (cf. Section 2.2).

The structured nature of the interviews is expected to achieve comparability between the respective applicants' individual performances. More so, the fact that an applicant's performance is rated per question and the questions are completed in a definite sequence provides additional fairness by ensuring that no questions are skipped or impromptu questions are added leading to all applicants having a chance at answering the same set of questions.

To enforce compliance with the time limits for both parts, a timer intended to remind the recruiters not to spend too much time on a single question runs during both phases, since all questions must be discussed and assessed before the personal interview. This measure aims to limit certain biases that could arise due to the content of the chat conversation. For example, if the chat slips into private territory because the recruiter and applicant happen to share similar interests. The mandatory sequential handling of the questions can ensure, at least, that all questions have to be discussed before the conversation ends.

Anonymity The anonymization of applicants as well as recruiters is designed to eliminate a number of unconscious biases that are typically tied to a cognitive or perceptual trigger. Getting rid of those triggers through anonymization provides recruiters with the opportunity to receive a first impression of the applicants independent of their potential implicit biases. Applicants' personal information, however irrelevant it might be to their job performance, impacts their chances of success when it comes to hiring decisions. An applicant's name, accent, or skin color can activate (unconscious) biases in a recruiter. Therefore, hiding this information is a crucial aspect of the system. This is addressed by conducting the job interviews anonymously via asynchronous text messaging. This feature is designed to eliminate the biases typically present in classic job interview settings. No personal information about the candidates is disclosed to the companies beforehand.

Applicants as well as company representatives are referred to only by aliases so that applicants' identities can remain hidden from the companies until they are revealed in the face-to-face interview.

While the face-to-face interview may seem to undermine the anonymization effort, it was integrated into the process as a concession to company representatives who had uttered the need to see a candidate to be able to get an idea of who they are in the pre-study. This aspect was identified to be critical to a positive acceptance of the tool. It also makes the sequence of the interview phases all the more important. To minimize the possible impact of bias that might result from the de-anonymization in the face-to-face interview, recruiters must have no insight into the participants' looks, their habitus, body language, or any other personal information that might lead to a bias activation before they have sequentially gone through a pre-defined set of questions which can be comparatively evaluated. Conversely, the impressions recruiters gain during the chat interview and the resulting ratings are immutable. Cases showing big discrepancies between the impressions gathered during the structured and the unstructured interview can promote a reflection process by making this contrast visible to company representatives.

While numerically both phases of the interview receive the same weight for the calculation of the applicant ranking, the chat interview receives an implicit emphasis by allocating twice as much time to the first interview phase. While the textual exchange makes up 30 minutes of an interview slot, the interview done in person is scheduled to last no longer than 15 minutes. The normative emphasis that is put on the structured part of the process also reflects the relevance of the evaluations: While each question in the chat is available as an individual item in the overall applicant evaluation, the personal conversation is only represented with a single item, i.e., in the context of the complete evaluation, it is only one of 7 rated items.

5.4 Core Functionality

The underlying core concepts translated into a set of features making up the core functionality of the software tool that was to be developed. Three user roles had to be present in the system: The *administrator* user role is responsible for the overall setup and maintenance of the system. Tasks performed by *administrators* include creating the user accounts and populating the database with the company data, such as questionnaires and interview pairings. Company representatives, who conduct the job interviews and rate applicants' performances, are represented in the system by the *recruiter* role, while students applying for a job are represented by the *applicant* role. In the following I will outline the steps of the core process, modeling the recruiting procedure for a specific job opening in a company.

1 Populating the Database

In the first step of the process, *administrators* populate the system database with company and applicant data. This encompasses the creation of user accounts for applicants and recruiters, as well as entering questionnaires for each distinct job opening. The questionnaires are defined by the companies who assign a weight to each question per how important they deem it to be in the context of the specific questionnaire. Company representatives receive training about the fundamental requirements the questions should meet as part of a dedicated workshop they participate in to prepare for the event.

2 Answering Questionnaires

The second step involves *applicants* answering questionnaires assigned to them in writing. Applicants are aware of the nature of the job offerings and the specific companies they apply to when answering the questionnaires. They are advised to keep their answers concise to keep the time necessary to read and comprehend the answers during the next step in the process to a minimum. This helps to ensure company representatives do not take too much time to read the prepared answers before they can ask contextual follow-up questions so that there is enough time for meaningful discussion. *Applicants* answer the questionnaires asynchronously on their own time. The questions are not time pressured so they can take however long they feel is necessary to make their answers as concise and meaningful as possible. To be able to proceed with the next step, an *applicant* must have answered all questions of a questionnaire. Only then can they proceed to the job interview.

3 Conducting Job Interviews

The third step is the only part of the process that functions synchronously. *Recruiters* can see which *applicants* have completed all questions of their questionnaire and invite them for the job interview when they are online. Once an *applicant* has accepted said invitation, the two-tiered interview process can commence. In the first phase a textual interchange, structured by the sequential order of questions and the *applicants'* corresponding answers, helps recruiters to gain an impression of the candidates without it being tainted by common biases. The applicant's answers are revealed to the recruiters one by one, each functioning as the basis for further discussion. When company representatives think that an answer was sufficiently discussed, they score the applicant's performance in the context of the specific question and can then proceed to the next question. Once this procedure has been completed for the whole questionnaire, they proceed to the second phase: an unstructured face-to-face interview.

4 Rating and Ranking Applicants

After the face-to-face interview is completed, recruiters can add a textual evaluation and a numeric score for the interview performance marking the final step of the interview process per candidate. Once all applicants for a job opening have been interviewed,

recruiters rank applicants based on their job interview performances. The tool offers reporting functionality in the form of a comparative, sort-able table view, allowing the juxtaposition of applicant scores for each question to support the ranking process.

5.5 Participant Workshops

Two workshops aimed at raising awareness for unconscious biases were held in preparation for the upcoming recruiting event, facilitated by a diversity consultant, with the goal of helping individuals understand and recognize the biases they may hold, and providing potential strategies for addressing them on a fundamental level.

The workshops were devised by the *TU Career Center* to serve different purposes depending on the audience group. For company representatives, the workshop was intended to garner acceptance for the format by educating people on unconscious bias and confronting them with their own potential biases.

Meanwhile for the students, although covering much of the same subject matter, the workshop was intended to demonstrate how their chances of being hired or even invited for an interview might be affected by unconscious biases.

The workshop began with a brief overview of the concept of unconscious bias, defined as mental shortcuts that individuals use to process information, which can result in biased decisions and behaviors. The facilitator provided examples of common biases that people may hold, such as ageism, sexism, and racism, and explained how these biases can influence the hiring process and impact the diversity of a workplace.

To help participants understand the importance of diversity and inclusion in creating a positive and equitable work environment, the facilitator provided exercises aimed at helping participants identify and recognize their own biases and consider ways in which they can actively combat these biases in their personal and professional lives.

To ensure the smooth operation of the DEBIAS process, the workshops addressed two key areas. First, students learned strategies to answer questionnaires without unintentionally revealing details that could trigger interviewer bias. This was intended to help them participate in the DEBIAS process while protecting their anonymity. Second, companies received guidance on crafting questionnaires and avoiding specific follow-up questions during the chat interview, a measure aimed at fostering a more neutral and objective interaction between students and company representatives.

The workshop concluded with a technical introduction to the DEBIAS tool. Participants were explained the interview process including the filling out of the company questionnaires and how they were expected to interact with the software in the course of the event. Please refer to Appendix I for the presentation slides of the technical introduction in their entirety.

Part 3

Implementation & Evaluation

CHAPTER 6

First Iteration

In the following chapter, I will describe the user interface design, implementation, and evaluation of the DEBIAS software tool which aims to provide an equitable job interview experience based on the core principles and requirements established in the pre-study. The user interface plays a crucial role in the overall effectiveness as a means of promoting equitable job interviews. The design of the UI must not only be intuitive and easy to use for both applicants and staff involved in the recruiting process, but it must also effectively implement anonymization techniques to obscure identifying information about candidates. In the first section of this chapter, I will delve into the specifics of the UI design and implementation of DEBIAS, starting with an overview of the technologies employed in the system. I will discuss how these technologies were used and what purpose they served in the overall design of the tool.

The second section of the chapter will focus on the design of the UI itself. This will include a detailed description of the various screens and features of the DEBIAS software, as well as the thought process behind their design and implementation. I will also discuss how the UI was designed to effectively integrate anonymization techniques, and how it was tailored to meet the core principles and requirements established in the pre-study.

Finally, the third section of the chapter addresses the evaluation of the DEBIAS system in the form of user tests, by describing the user testing methodology employed as well as a presentation of the evaluation results. This includes a brief discussion of the results of the user tests and how they informed the ongoing development of DEBIAS. Research question RQ3 will be addressed in in Section 6.3.3 while Section 6.3.4 is concerned with answering research question RQ4.

6.1 Technologies and Development

DEBIAS was developed as a web application. This decision was driven by the need for students to commit their answers in preparation for the event and the necessity to provide access to the software on a minimum of thirty devices, which would otherwise have required an elaborate deployment scheme. Given my prior experience with the Python programming language and its versatility, as well as the availability of web development frameworks for the technology, it became the weapon of choice for the implementation of DEBIAS. In the following section, I will provide a brief overview of the technologies and software components used in the development of the DEBIAS software.

5 Backend and User Interface

For the application backend and the rendering of the main components of the user interface, I chose to use the Python web framework *Pyramid* [116]. Its developers describe it as a goldilocks solution positioned between microframeworks and mega frameworks providing developers with enough functionality to complete any size project but not making decisions for them. As such it allows developers to add only those components required to achieve a given goal and gives them the ability to pick and choose as necessary. *Pyramid* includes a scaffolding engine that can generate project templates for a variety of purposes. These so-called *cookie cutters* provide a basic structure for a project that acts as a starting point and can be filled and extended by the developers as necessary in the course of development. The DEBIAS tool was based on a scaffolding template incorporating *SQLAlchemy* [117] for data persistence and *Jinja2* [118] as a template engine.

SQLAlchemy is an open-source ORM framework and SQL toolkit that allows for the efficient persistence of data in relational databases employing an approach based on the data mapper pattern. An ORM (Object-Relational Mapping) framework is a tool that enables developers to interact with a database using object-oriented programming concepts. Using an ORM framework can provide a number of benefits when developing a web application, one of which is the ability to abstract away the underlying database structure and database interactions, allowing developers to focus on the business logic of the application rather than the specifics of interacting with the database.

SQLite [119] was selected as the underlying database for the project, since it is a widely used database engine that is well-suited for projects with a limited number of concurrent users, due to its reliability and self-contained nature. Its small overall footprint in terms of both disk space and memory usage and the fact that it does not require a separate server process to be running to access the data came as additional benefits to the setup and maintenance of the system.

The *Jinja2* template engine was chosen for generating the markup of the user interface. It is a powerful tool for the development of web applications due to its ability to provide a flexible and efficient means of generating dynamic HTML content. One of the key benefits

of using Jinja2 is its ability to support template inheritance, which allows for creating a base template that other templates can then extend. This helps to reduce duplication of code and enables developers to more easily maintain and update their web applications, as changes to the template can be made in a single location, rather than having to be manually updated in multiple files. In addition to template inheritance, Jinja2 also supports the use of variables, loops, and conditional statements within templates, allowing for easily incorporating dynamic content into web applications.

Being one of the most popular web template engines for the *Python* programming language [120], *Jinja2* can be used to generate markup as well as source code. It is easy to debug, and, through automatic HTML escaping, features a built-in measure against cross-site scripting attacks.

Using a template engine provides several benefits for development. One of the primary advantages is the ability to separate the presentation layer of the application from the business logic. This separation of concerns helps to improve the maintainability and scalability of the application, as changes to the presentation layer can be made without affecting the underlying business logic. Template engines also allow developers to create reusable templates, which can be easily extended or customized for different purposes. This helps to reduce duplication of code and enables developers to manage the development process more efficiently.

The user interface itself was based on the open-source CSS framework *Bootstrap 4* [121] that contains design templates for typography, navigation, and many other interface components based on a combination of HTML, CSS, and optionally JavaScript. *Bootstrap* is aimed at simplifying the development of web pages by providing a set of colors, sizes, fonts, and layout options that can easily be added and applied to a project, resulting in a uniform appearance of all supported user interface elements across different web browsers.

Bootstrap is designed to be easy to use and customize, with a wide range of documentation and resources available to help developers get up and running quickly, making it a popular choice for web developers. Some features, such as intuitive navigation menus, custom form elements, and clear and concise typography help to create a usable user interface. Furthermore, *Bootstrap* offers a set of JavaScript components providing additional UI elements such as dialog boxes, tooltips, progress bars, and navigation drop-downs, all of which came in handy for the creation of the DEBIAS user interface.

6 Chat

The chat interview constitutes the main feature of the DEBIAS software. Therefore it was crucial to find a robust, stable, and reliable solution for the exchange of text messages that could be displayed within the web application. The web-based chat client *Converse.js* [122], a client-side implementation of the XMPP chat protocol [123] based on JavaScript, HTML, and CSS running locally in the web browser, met all these requirements. It can be integrated into practically any website, using one of three different display modes.

While the overlay mode displays each chat as its own little window on top of the website, the fullscreen mode reserves a whole browser tab exclusively for the chat client. Finally, the embedded mode embeds the chat client into an element in the DOM structure of the web page allowing for the seamless integration that was needed in the DEBIAS tool.

Converse.js is based on the *XMPP* (formerly Jabber) protocol, an open, decentralized, and extensible communication protocol for instant messaging, presence, and other real-time communication applications. It was originally designed to be used as a building block for messaging applications but has since been extended to support a wide range of use cases, including Internet of Things (IoT) applications, social networking, and collaborative software.

XMPP is based on the XML language and uses a client-server architecture, where clients (such as messaging applications) communicate with servers over a network connection. It supports multiple authentication mechanisms and provides a number of features to enable real-time communication, such as message delivery receipts, typing notifications, and the ability to send and receive messages while offline.

One of the key benefits of XMPP is its extensibility allowing developers to build custom functionality on top of the protocol by using a variety of extension elements and attributes. This has led to the development of a wide range of XMPP-based applications and services, including chatbots, gaming platforms, and real-time data distribution systems.

The XMPP implementation in *Converse.js* is based on *Strophe.js* [124], an *XMPP* library for JavaScript. Through a plugin infrastructure, developers gain access to the internal API, making it possible to, among others, programmatically send, receive, and manipulate XMPP messages and handle them accordingly. *XMPP* distinguishes between different message types defining the conversational context a message is received in. In total, the protocol supports five different types (cf. Table 6.1), two of which were particularly relevant for the implementation of the chat interview functionality. The conversational content of the interview was exchanged in messages with the type `chat` while the coordination of the interview process was achieved by modifying the way *Converse.js* handles `headline` messages.

By default, *Converse.js* simply displays messages with the type `headline` inside the current chat window, paying no regard to the content of the message. To be able to use `headline` type messages to coordinate the interview process between interviewers and interviewees, it was necessary to process the message content and distinguish regular `headline` messages from the messages generated inside the DEBIAS tool to coordinate the interview process and sync the interviewer with the interviewee side. Such messages would contain a message body with a single XML element `<debias>` with the two attributes `session`, and `cmd`. The `session` attribute contains a unique identifier that is generated by the DEBIAS system whenever a chat interview is initiated by a *recruiter*, serving as a reference to a specific chat session. This way it is possible to ensure that a `headline` message sent to a user is only displayed in the context of a specific chat session. The `cmd` attribute contains one of four values that are processed and

Type	Usage
`chat`	message sent in the context of a one-to-one chat conversation
`error`	an error has occurred related to a previous message
`groupchat`	message sent in the context of a multi-user chat environment
`headline`	message generated by an automated service that delivers or broadcasts content
`normal`	single message sent outside of a one-to-one conversation or group chat

Table 6.1: Message types of the XMPP standard according to RFC3921

interpreted to trigger the appropriate actions in the user interface. A `chat-started` message is sent to the respective *applicant* client when the *recruiter* clicks on the button *Besprechen*[1] to start the interview, after which the *applicant* is prompted to join the chat. Clicking the join button triggers a `chat-joined` message on the *applicant* side that, once received, enables the *recruiter* to click start the actual interview process by clicking the respective button in the UI. This then triggers the candidate's first answer to be revealed to the interviewers as well as it being shown within the chat window so that it appears at the appropriate location in the conversation. The candidate's pre-submitted answers are transmitted in the form of normal `chat` messages prefaced by the keyword */debias* followed by a JSON body containing the respective interview question and the corresponding answer in a structured form so that they can be distinguished by the other messages within the conversation by giving them special formatting. Whenever a *recruiter* proceeds to the next question, a `question-end` message is sent to indicate that the current question has been rated, followed by another `question-start` message if there any more questions included in the questionnaire or a `f2f-started` headline message to indicate that the chat interview has been completed and the in-person portion of the process can take place. At the very end of the in-person interview, once the interviewer has submitted their score for the face-to-face interview, the client on the applicant side receives a `interview-completed` message which, in case the conversation has taken place in the form of a video interview, will close the video session for the candidate.

7 Docker Setup

In recent years, *Docker* [125] and containerization have gained significant popularity as a means of deploying and managing applications in modern computing environments. Containerization involves the use of lightweight, standalone, and executable packages called containers that contain all the necessary code, libraries, and dependencies to run an application. These containers can be easily deployed and run on any host, regardless of the underlying operating system or infrastructure.

Docker is an open-source platform that provides tools and services for building, deploying, and managing containerized applications. It allows developers to package their applica-

[1] German for *discuss/talk about*

tions into standardized containers that can be easily shared and deployed across different environments.

The use of containers and containerization technologies like *Docker* can provide numerous benefits, including increased efficiency, portability, and scalability of applications. They can also help to ensure consistency and reproducibility in the development and deployment of applications, making it easier to manage and maintain complex environments.

In the context of an application consisting of a client and a server, *Docker* can facilitate the development and deployment process by providing a consistent environment for both the client and server components. For example, each component can be built and tested in its own container, which ensures that the same environment and dependencies are used throughout the development process. This can reduce the risk of issues arising due to differences in development environments and can make it easier to debug and troubleshoot issues. In addition to facilitating development, *Docker* can also make it easier to deploy the application in production. By packaging the client and server components into separate containers, they can be easily deployed and scaled independently of each other.

For the DEBIAS tool, the *Docker* setup consists of three containers linked together in a virtual network through docker-compose. In the context of Docker, a virtual network is a logical network created within a *Docker* engine that allows containers to communicate with each other and with external networks. Virtual networks can be used to isolate and segment traffic between containers, which can be useful for security and networking purposes.

The web application itself, including the database, is located in the first container, while the *Prosody XMPP* server [126] is hosted in a second container. Finally, the third container houses an instance of *NGINX* [127] that acts as a reverse proxy, meaning it receives requests from clients and forwards them to other servers or services. This can be used to distribute traffic across multiple servers, improve performance and scalability, and enable communication between different domains by bypassing the same-origin policy.

In the DEBIAS setup, the `proxy_pass` directive of *NGINX* was used to allow the instance of the modified *Converse.js* client to access the HTTP Bind endpoint of the Prosody server. HTTP Bind (also known as BOSH) is a protocol that allows a client to maintain a long-lived HTTP connection with a server in order to facilitate real-time communication. When a client connects to the server using *Converse.js*, it establishes an HTTP Bind connection by sending an HTTP request to the server at regular intervals. The server responds to these requests with a message acknowledging receipt of the request, allowing the client to maintain a persistent connection with the server and receive timely responses to its requests.

Employing the `proxy_pass` directive was necessary due to the same-origin policy of modern web browsers, a critical security mechanism that restricts how a document or script loaded from one origin can interact with a resource from another origin. This policy helps isolate potentially malicious documents, thereby reducing possible attack vectors. It prevents a script from making requests to a different domain than the one it came

from, which if allowed, could lead to web attacks like Cross-Site Request Forgery (CSRF), Cross-Site Script Inclusion (XSSI), and others. Without the use of the `proxy_pass` directive, the same-origin policy would have blocked the *Converse* client from directly communicating with the Prosody server, rendering direct communication impossible.

6.2 User Interface

While the login screen is the same for all users, once they are logged in, they have access to distinct sets of functionalities in the system, depending on their user role. *Applicants* can only answer questionnaires and accept interview requests, *recruiters* receive an overview of the interviews they will be conducting and the possibility to compare applicant ratings. In the following, I will provide an outline of the main points of interaction for each user role.

6.2.8 Applicants

The following section describes the main points of interaction with the system for *applicants*, including the typical workflow that needs to be completed, starting with answering the company questionnaire in preparation for the event, accepting a company's interview request, and finally the interview itself.

Home Screen

After logging on to the system, *applicants* are greeted with a home screen, displaying all relevant questionnaires for their assigned job openings in the form of an overview. By clicking the button *Beantworten*[2] the user can open the questionnaire to answer the individual questions. Users can fill out the questionnaires and edit their answers right up to the point when the interview process has started. Since the *recruiters* can not review their answers before the interview, it is not necessary to prohibit users from editing their answers. Once an interview is in progress, however, *applicants* can no longer make changes to their answers to the corresponding questionnaire.

Once an interview has been completed, the button *Beantworten* is replaced by *Ansehen*[3]. Clicking the *Ansehen* button allows users to review their answers, in case the respective interview is still in progress, and gives them the opportunity to review the whole textual exchange, including the scores they received from *recruiters*.

Answering a Questionnaire

When answering a questionnaire (c.f. Figure 6.1), *applicants* are presented with a card item for each individual question. The question itself is displayed at the top of the card. In case they have already answered a question, the user's answer is displayed beneath the

[2]German for *answer*
[3]German for *view*

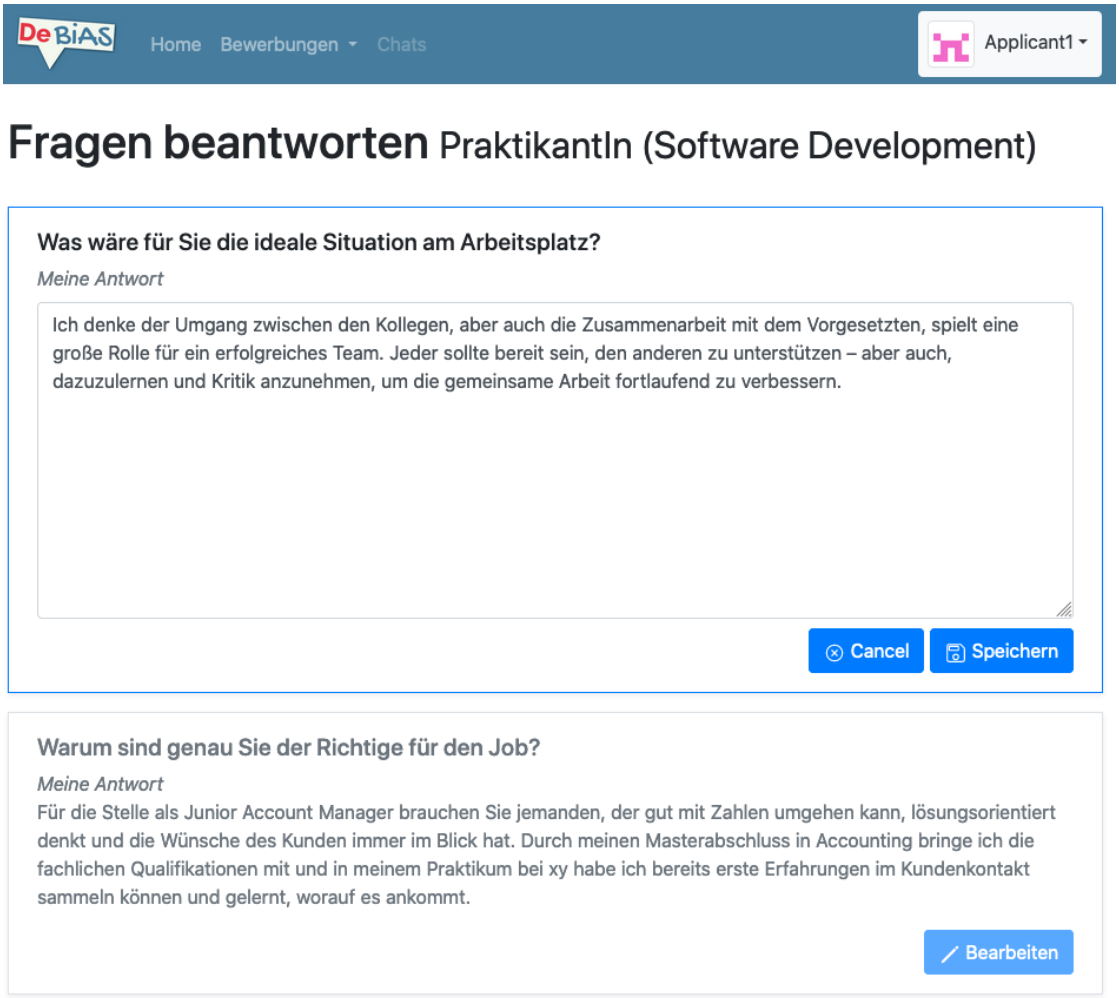


Figure 6.1: Questionnaire being answered by an applicant

respective question. If a question is yet to be answered, it can be opened for editing by clicking the *Beantworten*[4] button. This highlights the current card and adds an editable text area to it, while all remaining cards are being grayed out. The user can save their answer, once they are happy with it or discard it by clicking *Cancel*. If answering a completely unanswered question is canceled, the question will still register in the system as being unanswered, thus preventing the interview process from being started. Once an answer is saved, the question is counted as answered, however, it can still be edited until the interview process is initialized by the *recruiter*.

[4]German for *answer*

Interview

Once an *applicant* has accepted an interview invitation, they are redirected to the interview page. The layout here is kept intentionally simple. Most of the page real estate is taken up by the chat box used to communicate with the *recruiter*. Other than that, *applicants* can see the logo and name of the company they are currently communicating with in the page title. The first question and the *applicant's* respective answer are displayed inside the chat window when the *recruiter* has started the interview.

Top Menu

In addition to opening a questionnaire for answering through the home screen, *applicants* can also reach the same view via the top menu. Additionally, the option *Chats* displays chat sessions that were started but not completed. This allows users to rejoin an interview in case of technical difficulties that might result in them dropping out of the chat.

6.2.9 Recruiters

In the following, I will describe the DEBIAS tool user interface as it pertains to the usage by company representatives, including the UIs to conduct anonymous job interviews as well as any administrative functions needed to coordinate the interview process and finally, to compare applicant performances.

Home Screen

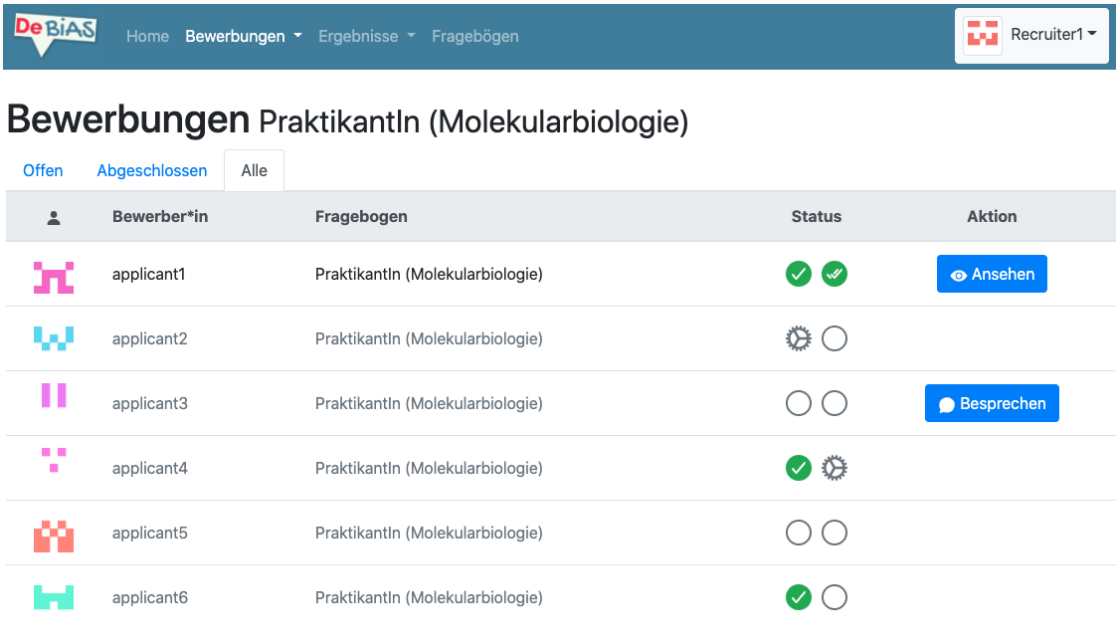


Figure 6.2: Home screen for the recruiter role with status indicators and action buttons

The home screen for *recruiters* (c.f. Figure 6.2) consists of a tabular overview of all applicants scheduled to be interviewed. The first two columns, starting from left to right, contain the applicants' aliases and their automatically generated avatar images. In the *Status* column, *recruiters* can see the current status of 2 stages in the interview process for each *applicant*, represented by two icons, the first of which indicates if the *applicant* has answered their questionnaire, while the second shows the status of the interview itself. Each icon can have 3 distinct states: an unfilled circle indicates that the respective stage of the process was not started, a cog indicates that the stage is currently in progress and a filled circle indicates that the stage has been completed. The user receives corresponding explanations of the indicators as tooltip-style popover messages when hovering their pointer over the symbols.

The column *Aktion*[5] contains one of two buttons, depending on what stages of the process have already been completed for the respective applicant. If the *applicant* has answered all questions of the questionnaire, the button *Besprechen*[6] is available to the user. The said button enables the *recruiter* to invite an *applicant* for a chat interview, ultimately initiating the chat conversation. Hence, the *Besprechen* button is grayed out until an *applicant* is online and therefore available to chat with. Once the interview process has been completed for an *applicant*, the button is replaced by the button *Ansehen*[7], which allows the *recruiter* to review the completed conversation including scores.

Interview

Since the *recruiters* are responsible for the administration of the individual interviews, the *recruiter* view of the interview process, as shown in Figure 6.3, is not quite as simplistic as the *applicant* version. Here the screen is horizontally divided, with the lower part being occupied by the chat box, while the upper part of the screen displays the current question with the *applicant's* answer alongside a couple of additional input elements. The top section functions as an indicator of the interviewing progress. The questionnaire is illustrated with a set of numbers, each representing one question in the questionnaire with the number of the current question being highlighted, so the user can grasp where in the questionnaire they are currently at. An additional progress bar functions as a graphic indicator for the interview progress.
Beneath the *applicant's* answer, a 6-level Likert scale allows the *recruiter* to rate the *applicant's* performance according to their valuation. The decision not to include a neutral value in the scale was made because research indicates that the sheer presence of a neutral value might lead to people picking the neutral option by default [128]. The *recruiter* proceeds to the next question by clicking the *Weiter*[8] button, which also persists the rating for the current question.
A countdown timer helps *recruiters* to keep track of time during the interview, displaying

[5] German for *action*
[6] German for *discuss/talk about*
[7] German for *view*
[8] German for *proceed*

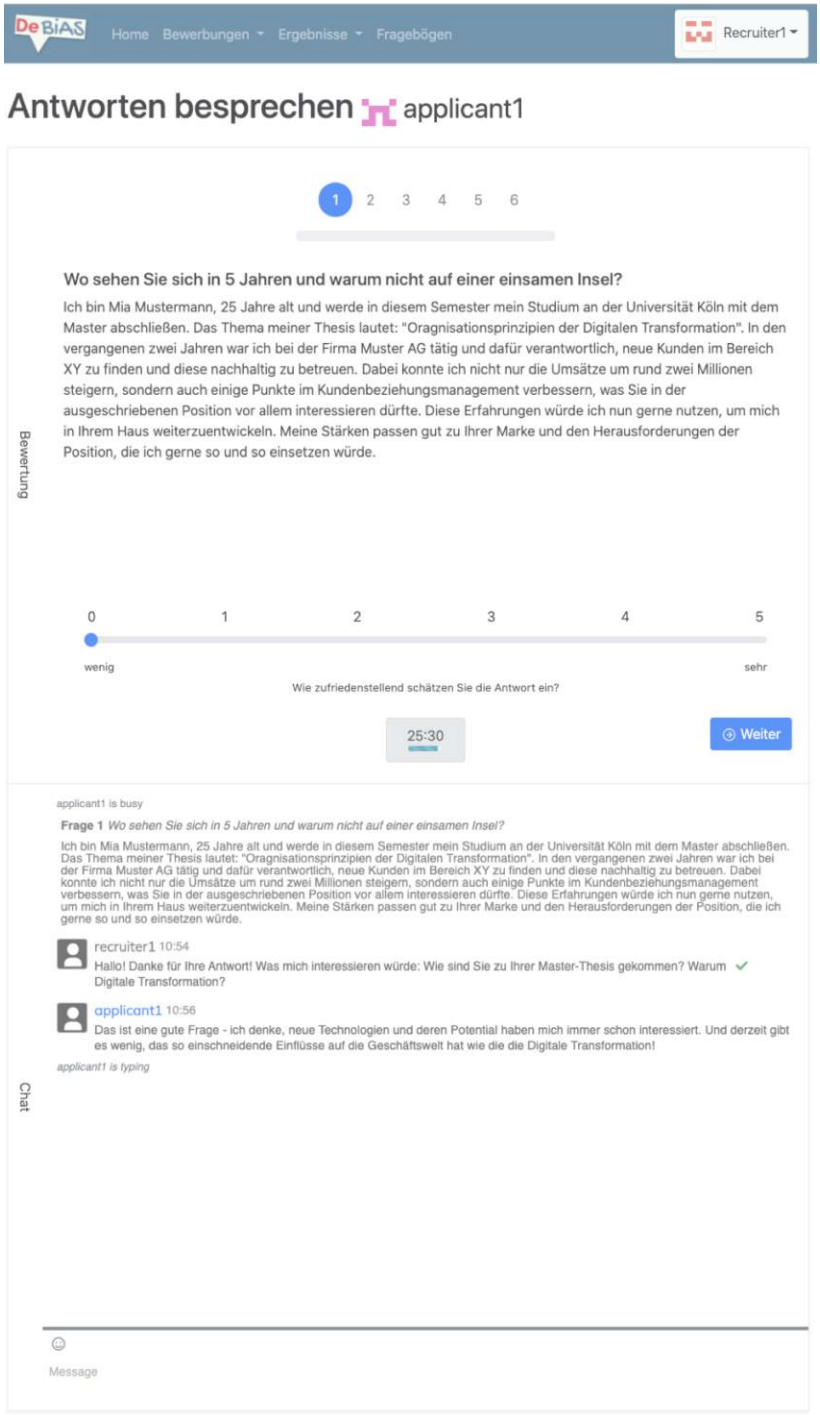


Figure 6.3: Interview page with question/answer pair, rating scale, timer, and chat box

the remaining time in minutes and seconds plus providing a graphic representation via a time bar, that shows how much time is still left proportionally. The timer changes its color scheme to yellow after half of the time is over and to red when there are less than 5 minutes left on the clock. The timer is purely informational and does not trigger any actions in the interview process when it runs out. It is the interviewers' responsibility to keep the interview on track and finish on time. However, if they should decide that they want to exceed the time limit, they are free to do so.
The chat box is fundamentally the same for *recruiters* as it is for *applicants*. In order to keep the experience consistent, questions and *applicants'* answers are displayed in the chat box as well. This is intended to orient the user by providing the context of the current conversation.

After the chat interview has been completed, the *recruiter* is presented with a simplified version of the user interface, that does not include the chat box, but adds a free-text field instead. This allows *recruiters* to add their notes and have them at hand at a later point when comparing *applicant* performances. Additionally, they rate the *applicant's* performance in the face-to-face interview on the same 6-level Likert scale used for assessing performance in the chat interview.

Comparing Applicant Performances

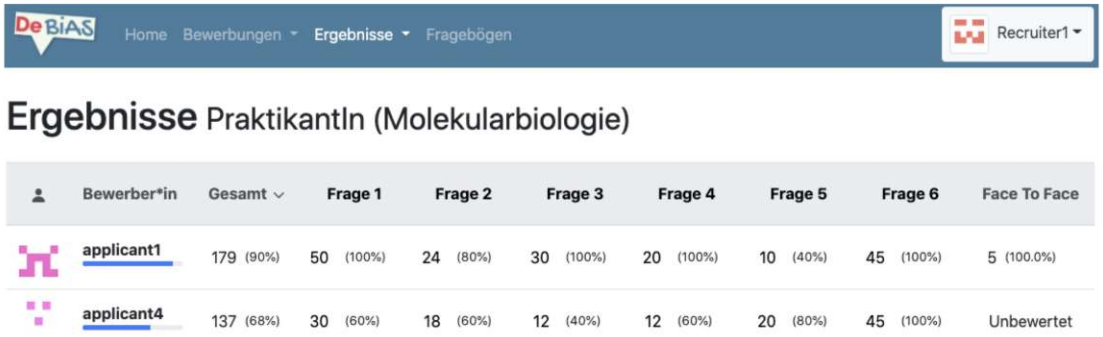


Figure 6.4: Tabular view for comparing applicant performances

Once all interviews for a job opening have been conducted, the tool helps to structure the decision-making process and determine which of the *applicants* to offer an open position. A tabular display shows not only the overall total scores received by the *applicants* but also the scores for every single question as well as the face-to-face interview, as shown in Figure 6.4. The scores are weighted according to the relative weight assigned to each

question by the company beforehand. An *applicant* may receive a maximum of 600 points for all questions in the textual chat combined and another 600 points for the personal interview. The actual number of points received for a specific question is proportional to the weight value assigned to it so that a question that was assigned a total weight of 50% results in *applicants* being able to get a maximum of 300 points for their answers to said question.
Additionally, *recruiters* are provided with a relative percentage indicating the relative amount of points the *applicant* has received in relation to the maximum possible score for the question. Alongside the *applicants'* aliases and avatar images, the interface provides a graphic representation of their relative scores in the form of a bar graph as a visual aid for decision-making. By clicking on the respective scores, *recruiters* can retrieve the whole conversation regarding the respective question in the context of the complete textual exchange.
All this is intended to mitigate cognitive biases associated with the volatility of our memory (cf. Section 2.2). By providing the performances of all *applicants* as a frame of reference and ensuring the granularity of the *recruiters'* assessments by providing the scores on a per-question basis, *recruiters* are prevented from focusing on candidates whose interviews were particularly memorable, especially for reasons not necessarily associated with their potential job performance. Especially anchoring effects, which might lead to one outstanding answer skewing the overall impression of a candidate, can be avoided by comparing them with the other candidates and reviewing their performances for the remaining questions.

6.3 Evaluation

The first iteration of the DEBIAS software was scheduled to be evaluated in the course of a recruiting event hosted by the *TU Career Center* that was conceived with the intention of providing a fair and unbiased recruiting experience to students. The evaluation of the system in a real-world setting allowed for data collection from both the people involved in making hiring decisions as well as the people applying for a job and undergoing job application processes.

6.3.1 Event Description

The six companies interviewed in the pre-study were also the ones to participate in the first event. Each company interviewed 18 of the 36 students. Consequently, each student interviewed with two companies. The event was set up to take place at *TU Wien* in two different, adjacent rooms. The first room called *Festsaal* was reserved for the company reps while the students were located in a second room called *Boecklsaal*. Figure 6.5 shows the floor plan of the two rooms. Eighteen students were present at a time, sitting at a desk with a laptop (cf. Figure 6.6). The company reps were either placed in interpreter translation booths in the next room (cf. Figure 6.6) or they had opted not to be on-site and therefore conducted the interviews remotely instead. Due to the rampant COVID-19

pandemic, two of the participating companies had instated policies preventing their employees from attending the event in person. For this reason, a video chat feature was added to the DEBIAS tool as a replacement for the in-person encounter. The company teams conducted a chat interview with each student before proceeding to meet them in person, or, in case they were not on site, in a video session, once the chat portion of the interview process had been completed. For the interviews that took place on-site, students would get up from their computers and physically change rooms to meet the company reps they had just been chatting with face-to-face. The event lasted the whole day with the first round of interviews starting at 9 a.m. and the last round ending at 6 p.m.

Participant Workshops

Leading up to the event, both participant groups took part in a workshop, as described in Section 5.5, aimed at making them aware of the potential influence of unconscious bias to provide them with context about the background of and reasons for the event. Participants would not only receive an introduction to the real-world consequences of unaddressed unconscious biases but would also be confronted with their own biases in a set of exercises guided by a professional consultant on diversity management and unconscious bias. Additionally, participants were introduced to the DEBIAS tool and how they were going to use it in preparation for as well as during the event.

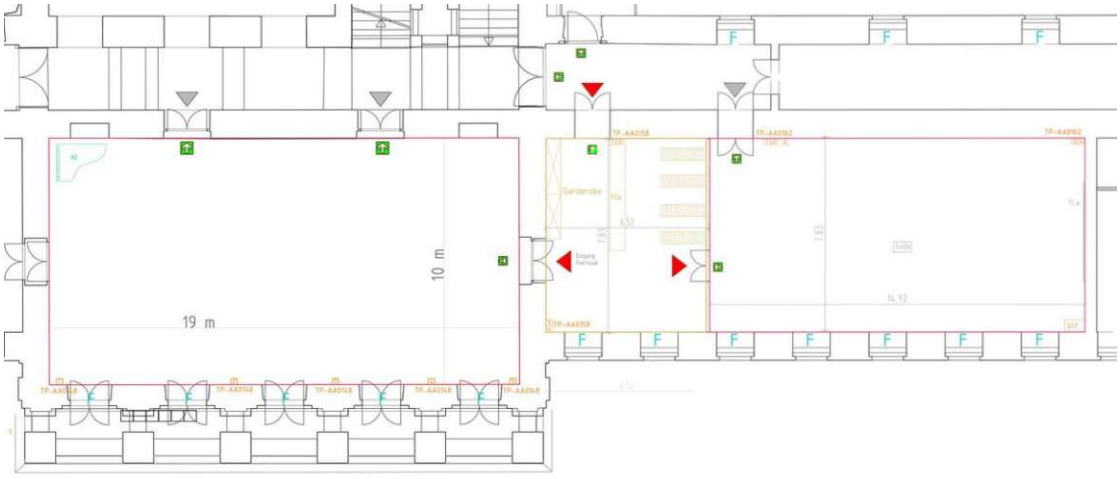


Figure 6.5: The room for the students (*Boecklsaal*, right) and the room for the company representatives (*Festsaal*, left) were located next to each other

Figure 6.6: Students in the process of interviewing with the companies via textual or video chat (left) and in person (right)

6.3.2 Methodology

Interviews

A set of short, semi-structured impulse interviews was conducted with the participants of the event. These interviews would take place right after the interviewee had completed at least one round of the interview process including both the chat interview as well as the face-to-face portion of the job interview. 22 out of 36 participating students were interviewed in the course of the first event. In addition, the company representatives conducting the job interviews on site were surveyed. Out of the four, two interviews were done as group interviews debriefing two representatives of the same company at once. Based on the interviewees' preferences, the interviews were conducted either in German or in English and lasted for roughly five minutes each. All participating individuals were asked to give their written consent by signing the consent form available in Appendix F. The interview guides used for the participating students as well as the company representatives are available in Appendix D and Appendix C respectively. All interviews were recorded using a digital voice recorder and were transferred to a personal computer for subsequent transcription. An orthographic approach was selected for transcription, excluding filler sounds and pauses to ensure legibility. Context information, such as if an interviewee was laughing when making a statement, was incorporated in the transcripts to help with later interpretation. In case an interview ended up including portions that were entirely off-topic and were deemed unrelated to the research topic said parts were left out of the transcripts leaving an indication that part of the conversation had been omitted. The finished transcripts were imported into *ATLAS.ti*, a qualitative data analysis software, to perform qualitative content analysis. With the help of the software, quotations were identified and marked in the collected material before assigning them with *codes*. The coding process was done in multiple iterations resulting in a total of 76 different codes in 331 quotations. The resulting codes were then subsumed under common umbrella terms, if applicable, and categorized into a set of larger themes, resulting in a structure of themes (represented in the software as code groups), codes, and sub-codes. For coding, the qualitative content analysis according to Mayring that had already proven helpful in the evaluation of the interview data from the pre-study (see Section 4.1.3 for

a more detailed description of the used methodology) was employed once again with the same coding scheme, namely defining a single word as the *coding unit*, a string of sentences in the context of the same question as the *context unit* and all things uttered by the interviewees in response to a single interview question as the *evaluation unit*.

Themes and Umbrella Terms Generally, the distinct codes were subsumed under larger descriptive themes to allow for a systematic evaluation of all statements talking about a specific topic. All things that were said about the chat interview, for instance, were labeled with the code *chat*. The next level of coding was usually prescriptive. Depending on if the subject had stated an opinion about an aspect of the topic, their statement was either categorized as *positive*, or *negative*. In case the subject had mentioned a specific aspect fitting under a larger umbrella term, said aspect would be coded as a subcode of the corresponding category. Taken together, codes contained three levels of hierarchy at most, two of which were usually descriptive. Statements talking positively about the chat in general would be labeled *chat: positive* while statements mentioning a specific aspect of the chat, ie. the fact that candidates got to know the interviewers before meeting them face-to-face, in a positive context would receive an additional level of code: *chat:positive:familiarization*. Statements that did not contain any value judgment and talked in a mere descriptive fashion about the topic at hand did not receive a specific prescriptive subcode. Rather, the subcode was categorized directly under the umbrella code: *chat:focus*.

Interviews with both students and company representatives were coded sharing a partially overlapping set of codes. For further analysis, however, the data sets were regarded as separate entities. This served as a precondition for being able to examine potential differences in opinion resulting from the diverging needs the actors taking part in the process might have.

Applicant Position Tables

Through company reps rating the applicants' performances as part of the DEBIAS process, both for the chat as well as the face-to-face portion, a set of data was available for analysis that could potentially provide insights into how applicants' perceived performances in the chat related to how they would be evaluated in the personal interview and finally how company reps would rank the candidates' overall performances. For this reason, the data was examined for indications of how the students' performances in the chat would bear on their perceived in-person interview and overall performance.

6.3.3 Analysis

For the analysis of interview data, the emerging themes are examined in the following, first from the perspective of the applicants and then from the company reps' perspective. This is intended to provide a balanced impression, highlighting different viewpoints and potentially diverging needs of applicants and company reps.

Chat Experience

The following section covers the participants' impressions of the chat interviews and their experiences. For the applicants, the focus lies on their self-presentation and their ability to represent themselves, whereas the company perspective is mainly concerned with the viability of the chat interview to form an impression of a candidate.

Applicants The majority of students reported having had a positive experience during the chat. While some mentioned being skeptical at first, participants largely felt that they could represent themselves well. One student even went so far as to claim that the chat allowed them to represent themselves "[...] even better than in a normal interview."[9]. This strongly suggests that the DEBIAS system provided the applicants with the means to present and express themselves to their own satisfaction. Interestingly, even students who indicated that their experience in the chat was not entirely positive usually still expressed that they were able to represent themselves well. While there was a small number of students saying they did not appreciate the chat portion of the process at all, the general consensus was favorable towards the two-step process. The following quote, uttered in response to being asked if they felt they were able to represent themselves well, nicely summarizes the prevalent feeling found in the interviews:

> *"To be honest: Yes. I didn't expect it to go this way but I feel like all three companies got an idea of who I was and what my skills were by the questions that they asked me in the chat."* [10]

Having the chat conversation as their first point of contact with the interviewers offered many students the opportunity to approach the interview in a setting that felt less tense to them than an in-person interview. When it comes to the chat interview, students did not report any instances of anxiety. In fact, they mentioned quite the opposite:

> *"I think that, among others, it is an advantage for me to have a kind of barrier which makes me less nervous than I would be if I had to present the sum total of myself - clothes and all."* [11]

The reduced tension the chat interview provided was one of the aspects that was appreciated most by the candidates. Many students mentioned that they felt quite relaxed during the chat interview which helped them ease into the process. Other than the anonymity the chat provided making them feel less exposed, students also acknowledged that the mode of communication gave them some extra time, allowing

[9] Quote by Applicant 06
[10] Quote by Applicant 09
[11] Translated from German by the author. Original quote by Applicant 08: "Ich denke, dass ein Vorteil für mich, unter anderem, ist, dass ich ein bisschen eine Barriere habe und deshalb nicht so nervös bin, als wenn ich mich jetzt persönlich mit allem – mit Kleidung und so weiter – präsentieren muss."

them to think more carefully about their answers. In consequence, they were more satisfied with the responses they gave:

> *"In the chat you have time to think about what you want to say and arrange your idea. So it would be better than just a normal interview."*[12]

All this led to applicants feeling more comfortable during the interview. Since they didn't expose themselves in the same way they would in an in-person encounter, they could not be judged for their appearances or slip up answering. The stress reduction also carried over to the in-person interviews. The chat helped to ease them into the face-to-face by allowing them to *"[...] break the ice [...]"*[13] and familiarize themselves with the interviewers, so that they didn't have the feeling of exposing themselves to a stranger. Instead, they faced a person they have already been in contact with:

> *"And then, after you get comfortable with the person you are talking to, you then go to the video chat. Then, simply be yourself because it is a continuation of what you have been discussing before."* [14]

Company Reps Overall, the interviewers expressed that the chat had worked well for them and that they were content with the execution as well as the substance of the chat interviews. Multiple company representatives mentioned that students seemed well-prepared and that the answers they received were well thought out. One interviewer explained this by students using the time in the chat to reflect on their answers more than would be possible in a face-to-face interview, *"[b]ecause there you can't let 40 seconds pass before giving an answer"*[15].

When it came to assessing the candidates' personalities, company representatives felt that the chat interview provided them with enough input to gain an adequate impression. To arrive at a conclusion about an applicant's personality, they did not exclusively rely on the pure text of the answers given by the interviewees. Rather, they took other information such as how long it took a candidate to reply, respectively how often they paused when writing their responses, and used that to support their assessments. This information was inferred by interviewers based on an indicator in the chat system informing the user if their vis-a-vis was currently typing. If a candidate paused frequently when crafting their answers this was interpreted as the question moving something inside the candidate and hence them thinking harder about their answer.

The crux of the project was to design a process that provides fairness through anonymization in job interviews while still meeting the needs of the people involved in making

[12] Quote by Applicant 24
[13] Complete quote by Applicant 24: "I think it's good to break the ice with the chat."
[14] Quote by Applicant 06
[15] Translated from German by the author. Original quote by a representative of company C6: "Denn da kann man keine 40 Sekunden verstreichen lassen bis man eine Antwort gibt."

hiring decisions. The linchpin to fulfilling this objective turned out to be the question of whether interviewers felt they were able to get an idea of an applicant's background and personality in the chat interview. Most company representatives reported they were able to form an image of who the person they were interviewing was. However, more often than not interviewers mentioned that they were surprised when they met the applicant in person or in the video chat, because their mental image of who they expected their vis-a-vis to be did not line up with the person sitting in front of them. A discrepancy between the interviewees' assumed and their actual gender was reported most frequently. This issue was even explicitly acknowledged by a company rep, who said that, when talking amongst them, they were implicitly referring to their interviewees by the pronouns *he/him* by default.

> *"It wasn't so much that we were surprised. But implicitly you always talk about him. Him. That's more or less how it goes in technology."* [16]

This implies that there is in fact the prevalent blueprint candidates are expected to conform to, consequently diminishing chances for applicants who do not fit this mold. This provides additional evidence for the necessity to address implicit biases present in selection interviews.

Despite interviewers reporting multiple instances of having been surprised by their vis-a-vis, the general consensus seemed to be that it was in fact possible to make an assessment of the candidates' personalities and professional areas of interest based on the chat conversation, albeit not in every single case. One recruiter mentioned that in the case of one candidate the impression they had gotten from the chat interview alone had been *"catastrophic"*[17]. According to them, the candidate exposed a strong focus on security and stability during the chat – a fact the recruiter interpreted as an indication that the interviewee showed little willingness to perform, instead seeking a secure position akin to a clerk. When the company rep finally learned that the candidate had only been living in the country for 3 years while struggling through, they were able to put their need for safety into perspective and hence evaluated it differently.

The above example shows how anonymization can cut both ways. While the mode of communication employed by this system prevents biases based on what the interviewers can see or hear, it is not able to prevent them from putting candidates into boxes based on whatever information is available to them.

[16] Translated from German by the author. Original quote by a representative of company C1: "Es war jetzt nicht, dass man überrascht war. Aber man redet halt implizit immer von *ihm. Ihm.* In der Technik ist es halt ein bisschen so."

[17] Translated from German by the author. Original quote by a representative of company C6: "Also ohne dem nachfolgenden Gespräch wäre mein Bild ein katastrophales gewesen."

Missing Feedback

In the following, I will explore the impact encountered by participants due to the lack of nonverbal cues in the conversations, a topic that appeared far more relevant to the applicants' than the company reps's experiences.

Applicants Students reported occasionally being unsure of how their answers were received by their vis-a-vis with some applicants finding it difficult to interpret the follow-up questions they received by the companies in the absence of nonverbal cues. Random instances of miscommunication were reported by the applicants, which they attributed to the lack of feedback, as is exemplified in the following quote:

> *"Sometimes it is fairly difficult to interpret the intention behind the questions coming from the companies. Simply, because you make little use of smileys and such. That is to say, answering on spec and having misunderstood the question did occur."* [18]

This apparent lack of contextual information was mitigated by some interviewers providing the applicants with explicit verbal feedback, as is exemplified by an applicant in the following quote:

> *"In the chat, it was pretty clear as well how the other person was reacting to what you were saying, because they would say comments like 'Okay, good answer!' or something like this."* [19]

Alternatively, the use of emojis served to enrich the textual exchange by providing the applicants with complementary information, thus supporting text interpretation. According to the students, interviewers seemed to use emojis very sparsely, but when they did it was usually much appreciated. While a lack of emojis seemingly gave the interviews an air of rigidity and formality, candidates reported that the interviews felt friendlier and gave them a better feeling if their vis-a-vis used emojis. The following quote aptly summarizes the phenomenon:

> *"The first one was more of a stricter... more traditional let's say, but both of them I enjoyed a lot. I mean even after going to the video chat I got to learn it was not really strict. Just the fact that they didn't use, for example, smileys made it seem like that."* [20]

[18] Translated from German by the author. Original quote by Applicant 05: "Es ist relativ schwierig manchmal zu interpretieren wie Fragen gemeint sind von den Unternehmen. Einfach weil man wenig Smileys oder so auch verwendet. Das heißt auf gut Glück dann zu antworten und die Frage dann missverstanden zu haben ist auch vorgekommen."

[19] Quote by Applicant 06

[20] Quote by Applicant 06

Appropriately, students described conversations where the interviewers did not use emojis as more formal as well as strenuous, whereas receiving emojis made them feel comfortable and confident:

> *"I received a couple of emojis, and I thought: now the company is satisfied and in a good mood"* [21]

A few students mentioned missing the human side in the interviews or feeling like they were not able to get their personalities across as well in the chat. Contrary to that, other students explicitly mentioned the presence of personal aspects and that they thought that the chat was more suitable for assessing someone's personality than their technical skills. This apparent discrepancy can possibly be explained by the differing focus in the distinct questionnaires: one of the companies, for example, said that they had put the focus on assessing the applicants' personalities in the chat interview while leaving technical questions to be addressed in the face-to-face part. Hence, one could argue that this could be as much a matter of interview content as of the mode of communication.

Company Reps The lack of feedback did not come up in any of the interviews with the company representatives. In regards to emojis, one person mentioned that they had noticed that the tool offered functionality to add them but they were unsure if that would actually work since they had not tried it. The same person uttered a request to include functionality allowing users to *like* a specific message akin to what can be found in social media or other text messaging tools.

Anonymization

The following section deals with anonymization as a central aspect of the process, examining weak points in its current implementation exposed by accounts of both applicants as well as company representatives.

Applicants As a central feature of the DEBIAS process, the applicants' anonymity is a prerequisite for a fair evaluation. The majority of students appreciated not having to expose themselves outright and having the opportunity to present themselves without their appearances having an influence. There were, however, reports highlighting potential limitations to anonymization in this specific context. One student mentioned, that the company representatives had correctly inferred her gender from the way she verbally expressed herself. The respective participant reported to have been identified correctly as a woman in two distinct instances, by two different companies:

[21] Translated from German by the author. Original quote by Applicant 19: "Also ich habe ein paar Emojis gekriegt und ich dachte die Firma ist jetzt zufrieden oder in guter Laune."

> *"So, they assumed I was a woman – both companies – merely based on my specific choice of words."* [22]

A second student said that the company reps had straight out asked them where they were coming from. They understood the question to be posed without ill intent but expressed the feeling that the companies should have been more careful in their communication. In another instance, one student reported revealing their identity due to social conventions. The company representative interviewing them had introduced themselves by name so they felt it was only appropriate to reciprocate. The student acknowledged in the follow-up interview that they were not directly asked their name but they felt that they had to respond with their name to their vis-a-vis' introduction, since they had also mentioned their name. Finally, one applicant expressed concerns about being judged based on the fact that they were not fluent in German and hence opted to conduct the interview in English.

Company Reps While candidates reported instances of company representatives potentially compromising anonymization by not paying attention to the implications of their own communication, company reps only mentioned a single case of an interviewee giving up parts of their identity unprompted. The candidate disclosed their gender by using gendered language when describing their position in their volunteer activities.

> *"She outed herself in the chat now. In this case we already knew: okay, this one is going to be female."* [23]

While many of the problems exemplified above could be solved by instructing interviewees as well as interviewers to stick to a strict protocol, abstaining from any questions or answers potentially touching on the applicants' identities, it goes to show, that even a process specifically designed with bias elimination and anonymization in mind, still is dependent on the individuals involved, in order for it to work. Participants voluntarily giving up their identity to the company reps, is an aspect that cannot be accounted for. Appropriately, the existence of methods for undermining anonymization, such as gender prediction, highlights how detrimental the commitment by companies to an anonymous approach is for its success.

Time

In the following section, the impact of the relatively rigid time structure on the participants' experiences will be explored, examining the perceived effects of time pressure.

[22] Translated from German by the author. Original quote by Applicant 05: "Also, sie haben getippt, dass ich eine Frau bin – beide Unternehmen – und das nur auf Grund von gewisser Wortwahl."

[23] Translated from German by the author. Original quote by a representative of company C6: "Die hat sich jetzt selbst entlarvt im Chat. Da wussten wir dann schon: okay, die wird weiblich sein."

Applicants While many students mentioned that the chat gave them more time to consider their answers, thus resulting in them feeling less stressed, some also reported feeling time-pressed. This was mainly owed to two factors: A few students were worried about their typing speeds, especially if they didn't use their first language in the chat. This was exacerbated by the fact, that the laptops that the students were provided all came with German keyboard layouts. Students who were accustomed to a different keyboard layout reported to have struggled with speed when typing:

> *"I use a different, non-german keyboard. So this was a little slower for me."* [24]

The second factor contributing to applicants experiencing time pressure was the occasional lack of coordination between the two parties. Some students reported cases in which company representatives had already proceeded to the next question while they were still typing, trying to answer in the context of the previous question, so that, when they finally sent their response, it would appear in the wrong context:

> *"But for example, I already answered a question but there was this line with the next question and my answer came beneath that."*[25]

Company Reps Time was reported to be an issue by only one of the companies explicitly. The company representatives mentioned that the time went by incredibly fast and finally got tight during the chat interviews. One interviewer gave an example of how a combination of circumstances led to an especially tight time budget:

> *"I had the feeling, especially with the first applicant, that there were two aspects involved causing the long response times. Firstly, she is a person – judging from the way she presented herself – who is extremely accurate and doesn't want to make mistakes. And also, her English is not her greatest strength."* [26]

Software

The following section deals with the participants' experiences with the DEBIAS software, exploring the potential for improvement and possible pain points encountered by the users.

[24] Translated from German by the author. Original quote by Applicant 28: "Ich nutze ein anderes Keyboard, nicht das deutsche. Und das war etwas langsamer für mich."

[25] Translated from German by the author. Original quote by Applicant 28: "Aber ich habe zum Beispiel schon für eine Frage geantwortet, aber das war diese Linie und die nächsten Frage und meine Antwort kam danach."

[26] Translated from German by the author. Original quote by a representative of company C1: "Also ich habe, insbesondere bei der ersten Bewerberin, das Gefühl gehabt, dass 2 Aspekte mitgespielt haben, warum die Antwortzeiten so lange waren. Erstens ist eine Person – jetzt von allem wie sie sich gegeben hat – die extrem genau ist und keine Fehler machen will. Und wo Englisch nicht so eine riesen Stärke ist."

Applicants The overwhelming majority of students were happy with the tool and did not see the need for further improvement concerning the parts of the software they were using. They mentioned that the tool was structured clearly and complete in its capabilities. The simplicity of the user interface was appreciated and the fact that the chat interview was very similar to other typical chat interfaces contributed to the fact that applicants found it easy to use, as is exemplified in the following quote:

> *"It was very clearly structured. The introduction on Monday wouldn't have even been necessary. Everything was self-explanatory anyway."* [27]

A few potential improvements, however, did come up in the student interviews. Some of the students were irritated by missing manual line breaks in the answers, they had prepared in advance, when they were displayed in the chat window. Students had not seen the user interface for the company reps and hence were not aware that their answers were displayed to their vis-a-vis as they had entered them in a dedicated section of the UI. One student requested the integration of an auto-correct feature into the software. However, this feature can be provided by the browser used to access the application and was reported to have been used in this way by other students. So, it seems that not all of the provided devices had the feature enabled. Another suggested feature concerned progressing to the next question in the questionnaire. The interviewee proposed to add functionality that would require students to confirm that they were ready to proceed with the next item on the agenda. This feature could potentially avoid situations in which an applicant is still writing a statement in the context of a question when the interviewer has already progressed to the next question, as mentioned in Section 6.3.3. Finally, one applicant had a feature request concerning the video chat in the face-to-face portion of the process. They reported that a company representative wanted to send them a link during the interview, but they were unable to because the video interview did not have the possibility of a textual exchange.

Recruiters Similar to their student counterparts, company representatives were largely content with their experience using the DEBIAS tool. They found the tool "[...] self-explanatory and very easy [to use]." [28]. However, several suggestions and feature requests were made concerning the user interface of the chat interview. One company rep had run into a problem caused by the button used to proceed to the next question still having the focus after it had just been pressed. The respective interviewer described facing a situation in which pressing the return key by accident led to the software skipping to the next question without the opportunity to put in an evaluation score. Some of the interviewers found that the chat should have taken up more screen real estate. They considered it a waste of space that the current question as well as the applicant's prepared

[27] Translated from German by the author. Original quote by Applicant 29: "Es war total klar aufgebaut. Man hätte nicht einmal die Erklärung am Montag gebraucht. Das hat sich nämlich eh von selbst erklärt."

[28] Translated from German by the author. Original quote by a representative of company C6: "Sonst ist es selbst erklärend und sehr easy."

answer were displayed above the chat window as well as in the chat conversation itself. They expressed that they would have rather been able to see more of the proceeding conversation at all times. Other features requested by individual interviewers were a comment field that would allow them to take additional notes during the chat interview, as well as a "[...] dedicated timer for the current question [...]"[29]

6.3.4 Applicant Rank Data

Establishing comparability between candidates by having the company reps rate the candidates' performances on a per-question basis is a design measure intended to fight many of the cognitive biases that play a role in candidate evaluation (cf. Section 2.2). Based on the weight assigned to the respective question, each distinct rating factors into a total score that is made up of the combined weighted scores for the chat interview and the score for the face-to-face interview. Sorting the candidates by their total scores, the software provides the company reps with a ranking based on the points given to each candidate. However, at the end of the event, company reps were asked to rank all candidates they had interviewed by their own metric which might not necessarily reflect the assessment represented by the previously assigned scores.

To assess the validity of the automatically created system ranking, the manually assigned ranks were put in relation to the system ranks by calculating the absolute difference between the rank each applicant was given based on their overall score and their position in the manually created ranking. Ideally, both rankings would match, meaning that the assessment made in the DEBIAS process perfectly mirrors all criteria company reps use when making their hiring decisions. For the following evaluations, companies C1 and C6 could not be considered because the manually determined rankings of said institutions were not available to the author. In total, 36 students participated in the event, each interviewing with 2 companies resulting in 18 interviews conducted per company. It is important to note that the company reps were asked to provide a ranking without bindings. So, while multiple candidates might have received the same number of points in the DEBIAS process, therefore resulting in them tying for a position in the automated ranking, the company reps still needed to assign them a unique rank in the final position table.

Comparing the ranking based on the total score and the manually determined ranking, we find an average rank difference of 3.14, with individual per-company averages ranging from 2.78 to 3.33. These values indicate that the companies generally didn't diverge gravely from the assessments they made during the interview process, except for a few outliers.

The Spearman rank correlation coefficient [15] was selected as a metric to assess how predictive the system rank was of the final rank determined by a company. The said method enables the evaluation of a potential correlation between the rankings of two

[29]Translated from German by the author. Original quote by a representative of company C1: "Das eine wäre: es wäre cool, wenn ein eigener Timer für die gerade aktuelle Frage da wäre."

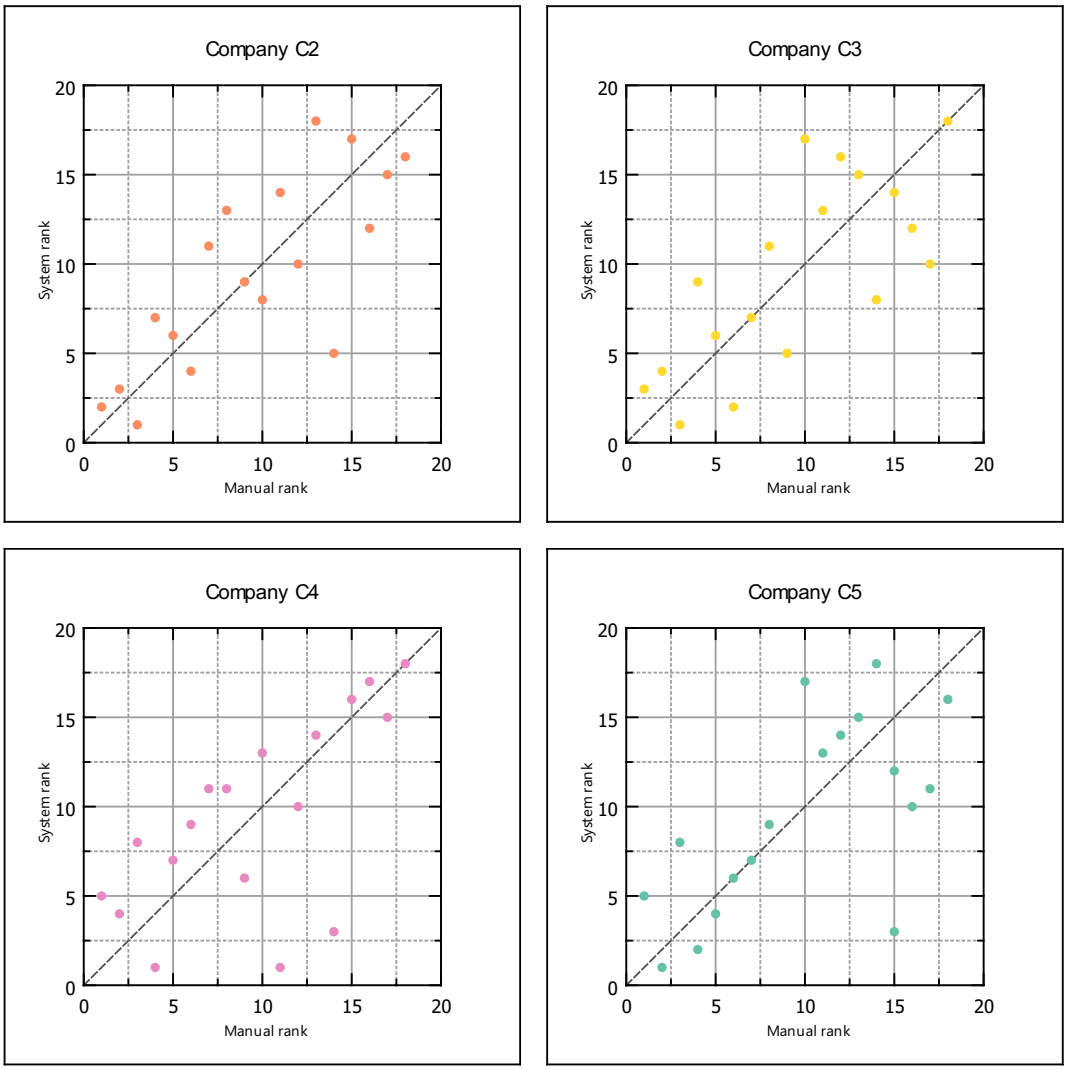


Figure 6.7: Plot of the applicant positions based on manual ranking as compared to ranked by overall score

variables. A Spearman correlation close to 1 indicates that observations have a similar rank between the two variables while a Spearman correlation close to -1 suggests an inverse relationship, i.e. observations ranking high based on the first variable rank low based on the second. A Spearman correlation close to 0 points toward no correlation between the two variables. The Spearman correlation was calculated on a per-company basis to evaluate the correlation between the manually determined applicant ranking and the system ranking based on the overall applicant scores as well as the correlations between chat performance, performance in the face-to-face, and the final assessment made

Company	C2	C3	C4	C5
Manual vs. System Ranks	0,781	0,738	0,652	0,609
Manual vs. Face-to-Face Rank	0,848	0,737	0,656	0,695
Manual vs. Chat Ranks	0,537	0,539	* 0,386	* 0,308
Chat vs. Face-to-Face Ranks	0,587	0,717	0,489	0,598

Table 6.2: Spearman correlation coefficients of applicant ranks for the first event

by company reps. The significance of the calculated correlation coefficient was tested by calculating the p-value. For the null hypothesis H_0, no significant correlation between the two variables was assumed. All hypothesis tests were conducted at a confidence level of 95% with a maximum p-value of .05. For p-values larger than .05, the null hypothesis was rejected. To assess the effect size, Spearman correlation coefficient values > 0.7 were considered to indicate strong, values between 0.3 and 0.7 were considered to indicate moderate, and values < 0.3 were considered to indicate weak correlations [129]. Table 6.2 summarizes the calculated Spearman correlations for said variables. Correlation coefficient values for which the null hypothesis could not be rejected were marked with an asterisk in the table.

Figure 6.7 shows a visualization of the candidates' manually determined versus their computed ranks based on their total scores for companies C2, C3, C4, and C5. The closer a data point lies to the diagonal, the smaller the difference between the manual and the computed rank. All points on the diagonal retained the rank the system assigned them in the manual ranking. The visual representation highlights candidates whose manually determined position diverges greatly from their system rank. Both companies C4 and C5 feature cases of applicants ranked ten positions lower in the manual than in the score-based rankings.

To answer research question RQ4, the rank correlations between the rankings based on applicant scores (system ranks) and the rankings manually determined by the company reps were calculated. For all companies, a positive correlation between the two variables was found with at least a moderate effect size with a strong correlation for two of them. This suggests that the results of the DEBIAS process, although they might ultimately be different from the final criteria employed by company representatives, indicate trends that can be found in the final decisions of the people involved in the hiring process and hence can function as a suitable indicator for what they consider good applicant performance.

The chat alone seemed to be far less indicative of the final position in the ranking than the ranks based on the combined scores of the chat and the face-to-face interview or even the face-to-face interview alone. For two companies a moderate correlation between the chat and the final ranks was found but for two more the null hypothesis could not be rejected based on the chosen significance level. Meanwhile, the face-to-face interview showed a correlation that was on par or even stronger than the chat and the face-to-face interview combined.

Comparing the ranks based on the chat interview with those based on the face-to-face conversation, the general trend pointed towards applicants receiving a great score in the chat also being ranked highly in the in-person interview. The correlation between the performance assessment in the chat and the in-person interview was found moderate for 3 out of the four companies and strong for the remaining one.

6.3.5 Findings

The evaluation showed that, while not everybody seemed to appreciate the chat interview the same way, most candidates not only felt confident in their abilities to represent themselves in the chat interviews but even seemed to profit from the two-tiered approach as it contributed to reducing their nervousness. For them, the chat helped to establish a certain level of familiarity that reduced the stress caused by being interviewed by a stranger.

Company reps similarly felt that they got a sufficient impression of the candidates in the chat – a fact that is also hinted at when looking at the applicant position tables. Examination of the Spearman rank correlation suggests a moderate to strong correlation between system rankings generated based on the performance values assigned to the candidates and manual rankings by the company reps for all participating companies.

When it comes to potential improvements, the time budget for the chat was mentioned by representatives of one company to have been too tight. A reduction in the number of questions per interview might improve this situation. Other than that, the fact that company reps skipped questions on accident because it was possible to proceed without explicitly giving a number rating presents further chance for improvement.

CHAPTER 7

Second Iteration

Building on the initial Debias tool evaluation, a second iteration was conducted with adjustments to the software, interview process, and evaluation methodology.

The adaptations made to the DEBIAS tool and the corresponding process were planned to be evaluated in the course of another recruiting event. The second installment was scheduled to take place 7 months after the first one but had to be delayed for an additional month due to uncertain attendance of students and companies being reluctant to participate given the escalating global pandemic. Consequently, the scope of the event was downscaled to only three participating companies compared to six in the first event.

The nominal number of applicants was reduced to 16, two of which cancelled the event resulting in a total of 14 students taking part in the evaluation. This allowed for a schedule that had each candidate interview with each participating company. Additionally, unlike in the first installment, all interviews were conducted on-site. The consent form all participants were asked to sign is available in Appendix G.

In the following chapter, I will first describe the changes that were made to the software as well as the interview process based on the findings of a preliminary evaluation of the data collected in the first event and explain the rationale behind each of the changes. In addition to the tweaks that were made in response to user feedback, I will describe a set of features that were added to make the system more robust and usable in a real-world scenario. A host of administrative functionalities was introduced to allow for more flexibility by allowing administrators to react to short-term changes more quickly and make the software more tolerant of user faults by giving administrators options to reset or correct faulty data.

The second section of the chapter describes the evaluation of the second iteration, detailing the adaptations to the methodology as compared to the first iteration. Evaluation results will be discussed in relation to the adaptations made to the user interface that were made in response to the feedback received concerning the first prototype. Finally, I will

discuss the impact of the two-tiered anonymous interview process as it presents itself in the collected data, addressing research question RQ3 in Section 7.2.1 and research question RQ4 in Section 7.2.2.

7.1 Adaptations to the User Interface

A preliminary evaluation of the data collected about the first iteration of the prototype served as input for further refinements of the software as well as the process. One of the main issues identified with the process was the tight time budget for the chat interview (cf. Section 6.3.3). Reducing the number of questions, as had been proposed by one company rep, promised to be a quick and easy solution. The questionnaires consisted of only four instead of five questions, in the second iteration.

Another issue encountered by company reps in the first iteration was the fact that pressing the return key would skip to the next interview question, under certain circumstances, without allowing the interviewer to rate the applicant's performance (cf. Section 6.3.3). This was addressed by disabling the button used to proceed to the next question in the interview, by default, only enabling it once the user had actually interacted with the UI element used to set a score for the performance rating.

Other than that, the user interface for the participants was only slightly refined, adding icons to some of the buttons or menu items and improving the integration of time slots into the overview for interviewers. Larger scale adaptations, such as complete layout changes, as proposed by representatives of one company (cf. Section 6.3.3), were not considered for the second iteration of the prototype to minimize the risk of breaking changes.

The bulk of the effort was put into further stabilization of the software and the implementation of administrative functions, allowing administrators to make changes to the software configuration on a running system, whereas there had been practically no way of adapting the configuration in the first prototype, once it was up and running.

7.1.1 Recruiters

Applicant Ranking

A single additional function was introduced in the *recruiter* user interface. At the end of each recruiting event by the *TU Career Center*, the companies are asked to provide a list of the students they have interviewed, ranked by the impression of their subjective performance. While this applicant ranking had only been submitted in paper form for all previous comparable events including the first iteration, the functionality to define and submit the company-specific applicant ranking was added to the DEBIAS tool in its second iteration. A tabular display shows the applicants' avatars alongside their aliases and displays their total scores and their per-question scores in absolute numbers, as well as relative percentage values in relation to how many points they could have gotten in

the best case. The relative total score is additionally displayed below the respective alias in the form of a bar graph as well, to allow for a quick comparison.

Recruiters start from the calculated applicant rank order, based on how applicants' answers were scored and what weight was assigned to the respective questions. They may then adapt the rank order of applicants by rearranging the table rows via drag-and-drop. Once the order has been adapted, an indicator shows the difference in position for each individual applicant, as compared to their calculated ranks. For applicants who were manually ranked higher than their calculated rank, the difference is displayed in a green badge with a leading plus sign, for applicants manually ranked lower, the difference is displayed in a red badge with a leading minus sign as shown in Figure 7.1.

Once they have concluded their ranking process, they submit their results by clicking the button *Abschicken* (submit) after which they are persisted in the database and can no longer be changed by the user.

Figure 7.1: Applicant ranking form with indicators showing the difference between computed and manually determined ranks

7.1.2 Administrative Functions

Configuration Import and Export

To allow for efficient backup and restoration of the system configuration a method of exporting the setup data as a JSON file was implemented. The export function, as shown in Figure 1 in Appendix A, has the option to include company logos as well as the automatically generated user avatars in the exported file. In case the user chooses not to include said data new avatars will be generated upon import of the file and the company logos need to be added manually later on. Since JSON is a human-readable format the export file can also be used to make adaptations to the configuration offline and feed them into the tool by uploading and importing an adapted JSON file. In addition to importing a complete configuration *administrators* can also import user accounts in the form of CSV files. The user first has to define which type of accounts they are going to import since there are varying requirements for the different account types. For instance, the data set for a *recruiter* account needs to include the respective company the account

is supposed to be assigned to. In case the file that is to be imported does not contain user passwords, or an *administrator* chooses to generate a new set of passwords, the toggle switch *Passwörter neu generieren*[1] can be turned on so that the import routine will automatically generate a passphrase for each imported account. Once a user import has been completed an adapted CSV containing the generated data alongside the input data provided by the user can be downloaded.

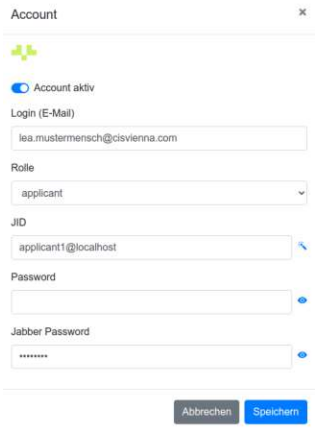


Figure 7.2: Modal dialog to add or edit a user account

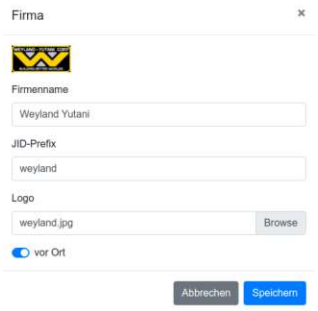


Figure 7.3: Modal dialog to add or edit company data

User Administration

Administrators receive an overview of all users as a tabular list, separated by user groups through tabs (cf. Figure 2 in Appendix A). Each line in the table contains the respective user's avatar, email address, and alias, as well as two buttons, the first of which can be used to reset the user's password and automatically generate a new one. This functionality was included to address users forgetting their passwords on location, in which case it is beneficial to be able to provide them with a new password quickly. The right of the two buttons opens a dialog (cf. Figure 7.2) enabling the user to view and change nearly all of the user account details for the respective user. From top to button, the dialog contains inputs to set the account status, the user's email address (*Login (E-Mail)*), the user role (*Role*), the user's alias in the form of their Jabber ID (*JID*), their password used to login into the application (*Password*), and the password for the chat server (*Jabber Password*), which is only used internally but has to be kept in sync between the chat server and the client application. For this reason, the *Jabber Password* can be looked up in the UI. The magic wand button next to the *JID* field automatically generates an alias for the applicant users by simply counting up the number of applicants. For user accounts

[1] German for *Generate new passwords*

that have been assigned the user role *recruiter*, an additional dropdown field is available, enabling the account to be associated with one of the companies in the database.

A series of buttons in the user overview provide extra functionality for user management, the leftmost (*plus*) of which is used to create a new user account by opening the same dialog used to edit an existing user account. Right next to it, the button with the *upload* icon enables *administrators* to upload and import a prepared list of user accounts as CSV files.
The button *Prosody vorbereiten*[2] starts a process that prepares all accounts on the chat server by adding all users they are scheduled to engage with in the chat to their contact list. This is an important prerequisite for ensuring that chat messages are not blocked by the system.
Finally, the button *Prosody Export* generates a list of commands to add all of the user accounts present in the database to the chat server via the prosody command line interface. The generated list can be downloaded for further processing by the user, for instance, to be used in a script setting up the *Prosody* installation.

Company Administration

Company Data Just like for the users, an overview of the companies is displayed as a tabular list providing *Administrators* with functionality to add as well as edit data for the participating companies. While the *plus* button in the top bar enables them to add a new company, the icon displayed to the right of each table row opens a dialog for editing an existing company entry (cf. Figure 7.3). The user can set the company name as well as the company logo. The given *JID-Prefix* is used when creating the user accounts for the representatives of the respective companies which receive a username consisting of a sequence number appended to the specified prefix.

Questionnaires The questionnaires the *applicants* need to answer are provided by the companies in advance and have to be imported into the system as CSV files. The administration interface provides functionality to upload said questionnaires and import them in a single-step process. Once the questionnaires have been added to the database, companies might still want to make adaptations. The *administrator* can incorporate said changes via the user interface shown in Figure 3 in Appendix A. They can change the order of the questions by rearranging them via drag-and-drop or change the relative weight of each question by setting a percentage value in the column *Gewichtung*[3]. The user interface also allows the addition of further questions, in case a company is unable to provide a finished questionnaire in time for the system setup. A total sum of all question weights is automatically calculated and provided to the user to facilitate the process. Additionally, changes to the questionnaire can only be saved if the weights add up to 100 percent. Each of the questions can be edited as well in a modal dialog that is opened by clicking on the edit icon at the right of each respective row. *Administrators*

[2]German for *Prepare Prosody*
[3]German for *weight*

can change the actual words of a question, both in English and German, and its assigned category (cf. Figure 4 in Appendix A). Each question is either categorized as personal or technical, represented in the overview by a heart (personal) or a hammer (technical) icon (cf. Figure 3 in Appendix A). The idea behind this categorization was to provide company reps with an additional analysis based on these aspects but due to time constraints, the feature could not be implemented for the second iteration.

Assignment of Interview Pairings

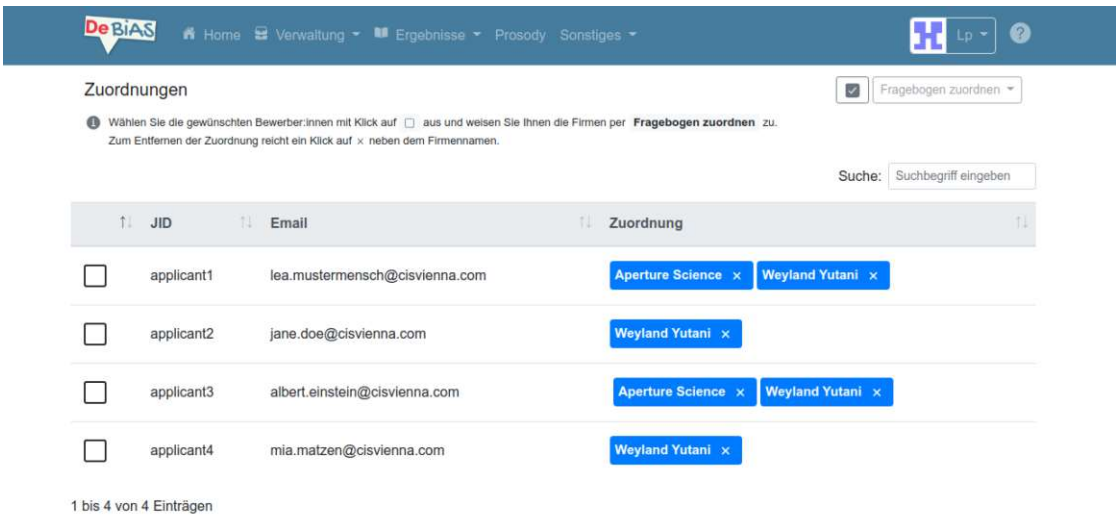


Figure 7.4: Bulk assignment of questionnaires to the applicants

Once users and companies including questionnaires have been set up, the *applicants* get assigned which companies they are scheduled to interview with. These assignments are typically imported from a CSV file together with the *applicant* user accounts but can be changed from the user interface if necessary to accommodate changes in schedule due to last-minute dropouts or special requests by *applicants* who might not want to interview with a specific company. A questionnaire can be assigned to one or more *applicants* at a time by first selecting the respective *applicants* in the tabular overview and then picking the respective questionnaire from the dropdown *Fragebogen zuordnen*[4] (cf. Figure 7.4). Assignments can be removed by clicking the *X* next to the company name in the respective *applicant* row.

Event Data Administration

Overview Other than setting up the system for the event, the most important task of an *administrator* in the DEBIAS environment is accompanying, respectively supervising the interview process to facilitate a smooth flow and ensure that the schedule can be kept.

[4]German for *assign questionnaire*

For this purpose, the conversation overview functions as a sort of dashboard displaying all scheduled interviews with their respective allotted timeslots. *Administrators* not only get a bird's-eye view of all interviews of the event but can also examine the status of every single interview. Even before the day of the event *administrators* need to keep track of each *applicant's* progress in answering the questions ensuring that the questionnaires are completed before the interview is scheduled to begin. During the event *administrators* keep an eye open for interviews that might run late, either because they start later than they were supposed to, or they go over time. The column *Gestartet*[5] shows the start time of an interview while the column *Bewertet*[6] shows the progress of an interview by displaying for how many of the interview questions a rating was submitted. Once a chat interview is completed, the column number of scored questions is replaced by a green checkmark icon. Finally, after the complete process including the face-to-face interview has been finished, the respective row is displayed with a light green background and the conversation protocols can be accessed by clicking on the document icon in the column *Konversation*[7] (cf. Figure 6 in Appendix A).

Ratings Since there had been cases of *recruiters* unintentionally saving applicant scores in the first iteration (cf. Section 6.3.3), functionality to correct applicant ratings was introduced. The decision to only provide *administrators* with the possibility to adapt the ratings after the fact was consciously made to prevent the data from being distorted. Having company reps go through an *administrator* to make adaptations was intended as a safeguard to ensure that changing applicant scores after an interview had already been completed remained the exception. Figure 5 in Appendix A shows an overview of all completed interviews grouped by company in tabs. The edit icon on the right opens the form shown in Figure 7 (Appendix A) to edit the score for a specific question while the trashcan icon allows the user to discard all scores for an interview, thus resetting the process so that it could be started from scratch again.

7.2 Evaluation

The data collected in the second installment were evaluated using the same methods that had already proven successful in the first event (cf. Section 6.3.2). This helped to establish a level of comparability allowing me to carve out the impact of specific changes made to the software in response to the feedback collected in the first evaluation. Consequently, for coding the collected interview data, the code structure established in the evaluation of the first iteration served as a baseline upon which then was expanded if necessary. Interviews were conducted with 13 out of the 14 participating students as well as with each of the three participating companies. While the interview guide for the company reps remained unchanged, the interview guide for the student interviews (cf.

[5] German for *started*
[6] German for *rated*
[7] German for *conversation*

Appendix E) was slightly adapted in response to some of the discoveries made in the evaluation of the first event.

7.2.1 Analysis

Chat Experience

Applicants The second round of interviews confirmed the findings from the first iteration with applicants largely reporting having had a positive experience in the chat interviews. Among the aspects appreciated most by the interviewees, the anonymity of the chat and the opportunity to think about their answers longer were brought up consistently. Once again, candidates mentioned that it helped them curb their anxiety to be able to familiarize themselves with the interviewers in the chat before they went on to talk to them in a face-to-face interview. The chat was perceived to provide a basis for the in-person conversation. This seems to confirm the observations made in the first iteration that getting to converse with the interviewers anonymously, can establish a level of familiarity, thus lowering the barrier of entry for the face-to-face, alleviating some of the anxiety of having to expose oneself to a complete stranger (cf. Section 6.3.3). The anonymity in the chat seemed to relieve applicants from some of the pressure of being judged. One participant even mentioned explicitly, that it had allowed them to express themselves more freely. When it came to representing themselves, students were overwhelmingly satisfied with sporadic reports of students struggling in a single one of their three interview sessions. The problems reported largely corresponded with what was found in the first iteration, namely some students having trouble using a keyboard layout, they were not accustomed to, and the chat feeling a little awkward. Just like in the first iteration, the occasional instance of miscommunication was reported by students. One interviewee made an observation about the differences in addressing miscommunication in the chat versus in an in-person conversation:

> *"Let's say, for example, you can stop me at any point to ask something in detail. I don't think that was possible with the chat. Because, like, I was typing like two, or three lines together. So, like, I had these small things, where I had to mention about, like, my previous studies, which was communication engineering, but they understood that as communications, which is like a totally different subject."*

The fact that your vis-a-vis typically receives one or more fully formed sentences – sometimes even a whole paragraph – in a textual exchange before they can react was recognized by the applicant to lending itself to talking on cross-purposes. In contrast, they are able to jump in and stop you at any point if they find what you are saying is not addressing the question when talking face-to-face.

Company Reps While overall, the people conducting the chat interviews were happy with the technical side of the process, some participants did express criticism about chat

interviews in general, finding the process to be quite time-consuming, with one interviewee stating that forming an impression of a candidate would have been much quicker if they had just talked to them in person. This was a sentiment that hadn't been expressed by any of the company reps in the first round of interviews. While the representatives of company C1, who had already participated in the first event, claimed that the impressions they had gotten in the chat were usually spot on – bar a single exception – for the other companies it was a different story. They reported the impressions they had gotten in the chat often being at odds with what they saw in the face-to-face interviews. One interviewee attributed this to the fact that they would supplement whatever information they did not receive about the person, effectively conjuring up an image of an applicant based on assumptions:

> *"You always form an image. That's just the way it is. But, due to the personal impression being missing fantasy supplements things depending on your expectations, doesn't it?"* [8]

Similar observations had been made by company reps in the first round of interviews (cf. Section 6.3.3). One company rep found that their ability or inability to assess a candidate based on the chat alone depended heavily on the quality of the candidates' answers:

> *"There's candidates giving direct answers, where you can really find out what you wanted to know. But then there's candidates who really answer something else. They want to describe it beautifully and I won't receive a straight answer."* [9]

It is potentially owed to this fact, that company reps reported often having had to revise their assessments made in the chat, once they had spoken to the applicants, finding that a candidate who first appeared to be very self-centered, presented as a great team player in the face-to-face interview, or a candidate whose answers seemed very aloof and matter-of-fact in the textual exchange turned out to be quite personable sitting across the table.
Asked about the viability of the process when it comes to reducing bias, especially representatives of company C3 were convinced that it worked. They acknowledged that

[8] Translated from German by the author. Original quote by a represenative of company C7: "Ja, man macht sich immer ein Bild. Das ist einfach so. Aber dadurch dass eben der persönliche Eindruck fehlt, ergänzt die Fantasie sozusagen je nach Erwartung vielleicht das dazu, ja?"

[9] Translated from German by the author. Original quote by a representative of company C3: "Es gibt Kandidaten, die wirklich konkrete Antworten geben und wo man wirklich herausfinden kann, was man will. Aber es gibt Kandidaten, die wirklich irgendwas anderes beantworten. Die wollen das schön beschreiben und dann kriege ich keine konkrete Antwort."

their usual hiring procedures were prone to biases, saying *"There's always sympathies or heuristics we are influenced by in deciding if we want to invite a candidate or not."* [10]

Anonymization

When asked about its capabilities to reduce bias, nearly all applicants were confident that the process was, in fact, able to provide a reduction in bias. One student, however, called the rationale behind the process into question, saying that it didn't make sense to them to do an anonymous interview first if their identity would then be revealed right afterward. Similar to what was reported about the first iteration (cf. Section 6.3.3) the second round of interviews revealed instances of stereotyping. Interestingly, one student reported having been assumed female due to their usage of gender-sensitive language. Cases of candidates outing themselves also occurred in both iterations. One of the candidates reported having revealed certain aspects of his identity, due to being asked to describe himself in a couple of sentences. According to him, his description of himself contained his gender, his age bracket, as well as his nationality, rendering the efforts to establish anonymity inherent in the process useless:

> *"The way I described myself definitely shows roughly who I am. I mean, the fact that I'm a man, my age, approximately. I explicitly wrote that I'm Austrian as well because that fits my description and then anonymization is gone, basically."* [11]

Said incident was also brought up by the company rep who had conducted the chat interview, remarking that applicants should be told not to reveal that kind of personal information.

Time

Applicants For most students, time appeared not to have been an issue. Similar to what I found in the first iteration, many students even expressed that the chat gave them more time to contemplate their answers than a conventional interview would have (cf. Section 6.3.3). Generally speaking, reducing the number of questions in the chat interview to four, seemed to contribute to the time budget feeling more adequate. One applicant, however, felt that time was still tight and they only were able to finish the chat interview on time, because they were well prepared. When asked if discussing four questions in 30 minutes was a manageable feat, they answered as follows:

[10] Translated from German by the author. Original quote by a representative of company C3: "Weil es gibt immer wieder Sympathien und Heuristiken, die uns beeinflussen bei der Entscheidung, ob wir einen Kandidaten einladen wollen oder nicht."

[11] Translated from German by the author. Original quote by Applicant 09: "So wie ich mich dann beschrieben habe, da kommt dann eindeutig raus so circa wer ich bin. Also, dass ich ein Mann bin, wie alt ich circa bin. Ich habe auch explizit geschrieben, dass ich Österreicher bin, weil das halt zu meiner Beschreibung von mir passt und dann ist diese Anonymisierung im Prinzip weg."

Company	C1	C3	C7
Manual vs. System ranks	0,927	0,937	0,760
Manual vs. Face-to-Face Ranks	0,870	0,867	0,700
Manual vs. Chat Ranks	* 0,264	0,881	0,554
Chat vs. Face-to-Face Ranks	* 0,029	0,697	0,719

Table 7.1: Spearman correlation coefficients of applicant ranks for the second event

> *"Yes, it was. As long as you had the answers already in your brain. If you had to think about them, you would definitely go over time."*[12]

Company Reps Although some company reps mentioned that the chat interview was a time-consuming process (cf. Section 7.2.1), interviewers did not report any issues with the time budget in the second interview round. A company rep who had already participated in the first installment of the event acknowledged the positive effect of reducing the number of questions in the chat interview down to four, saying having more time to spend on each question was a benefit. This suggests that covering four questions in a half-hour chat session is a manageable task and an additional reduction is not necessary.

7.2.2 Applicant Rank Data

While each candidate in the first iteration only interviewed with 2 companies, applicants would interview with all three participating companies in the second installment of the format. The three companies each conducted interviews with every participating student. Just like in the evaluation of the first event (cf. Section 6.3.4) the Spearman rank correlation was used to assess correlations between the rankings of applicants in the automatically generated position table and the position table created by the company reps manually. Once again, the null hypothesis H_0 assumes no significant correlation between the two variables, and hypothesis tests were conducted at a confidence level of 95% with a p-value of .05. Table 7.1 gives an overview of the calculated Spearman correlations. Correlation coefficient values for which the null hypothesis could not be rejected were once again marked with an asterisk. For the second event, position tables were available for all three participating companies.

When comparing the average rank difference between the manual ranking and the ranking compiled by the system the average rank difference over all companies was only half of what it had been in the first iteration (1.55 for the second vs. 3.14 for the first event) with company-specific averages ranging from 1.07 to 2.29. These values point towards candidates being ranked very closely to what their scores suggested. This interpretation is supported by the Spearman correlation as well. A strong correlation was found for all three companies with two companies even showing values over 0.9. When comparing the scatter graphs of applicants' manually determined vis-a-vis their automatically assigned

[12] Quote by Applicant 02

ranks for both events (cf. Figure 7.5) it becomes obvious that points are generally closer to the diagonal in the second iteration, meaning the difference between the automatically determined and the manually assigned ranks are smaller. Also, the number of applicants who received the same rank in both the system as well as the company position table more than doubled from 4 in the first iteration to 10 in the second.

The observation made in the first iteration, according to which there was a correlation between a candidate's performance in the chat and their face-to-face performance, held true for two of the companies. For the third company (C1) the variables appear to be more or less uncorrelated.

Examining the influence of the chat and the face-to-face portion on the final positions separately, no clear trend became apparent. While for company C3 the correlation between the chat and final ranks and the personal interview and the final ranks were more or less on par (0.88 for the chat vs. 0.87 for the face-to-face interview), the face-to-face conversation appeared to have weighed heavier for company C7 (0.55 for the chat vs. 0.7 for the face-to-face interview). Finally, for company C1 there was no correlation between the ranking based on the chat and the manually determined positions while there was between the personal interview and the final position.

While the overall correlations showed a greater effect size in the second generation, the ranking method employed by the company reps still appears somewhat unpredictable when focussing on some of the outliers. Like in the first iteration, some applicants who ranked in the bottom half of the candidate field for the manual ranking had received high scores both in the chat and the in-person interview. For a visual representation, examine the data points for companies C4 and C7 presented in the diagrams of Figure 7.5.

7.2.3 Findings

The evaluation of the second event iteration showed the same positive effects the two-tiered interview process had on many of the students in the first installment. Being able to acquaint themselves with their vis-a-vis while still retaining their anonymity made candidates feel less nervous when meeting the company reps in person. Once again, the majority of students were happy with how they were able to represent themselves and confident that the process helped to reduce biases.

While time pressure had been a prominent issue in the first iteration, addressing the problem by cutting down the company questionnaires to four questions seemed to have provided an effective solution. Neither students nor company reps expressed to have had an issue with the time budget the second time around. However, some of the company reps uttered critiques about the process of the chat interview being time-consuming in general.

Contrary to what seemed to be the prevailing sentiment in the first iteration, company reps mentioned the impressions they had gotten in the chat often being at odds with what they learned when talking face-to-face with a candidate. They attributed that

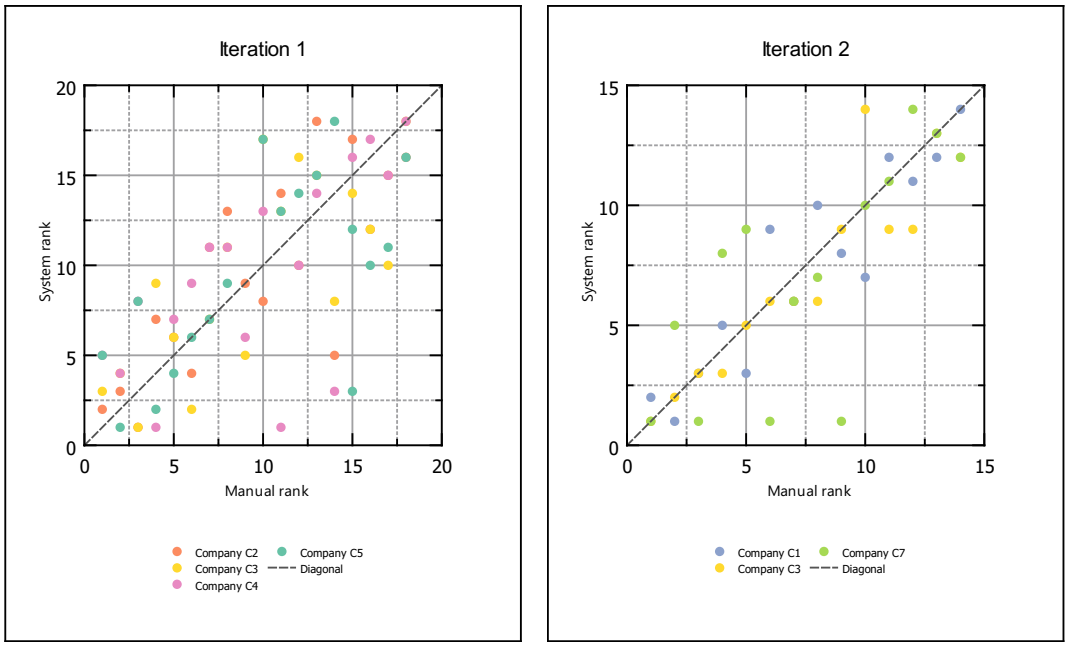


Figure 7.5: Scatter graphs for both iterations of the events comparing applicants' manually assigned ranks with their calculated ranks (system ranks) from all companies data was available for

to supplementing the limited information on the applicants by conjecture, therefore, conjuring up a flawed image of who they expected a candidate to be. This points towards the potential impact of biases even under conditions of anonymity. Interestingly, this phenomenon described by company reps was not reflected in the evaluation of the ranking data. While one would assume a smaller correlation between the ratings for the chat and the ratings for the in-person interview if the interviewers actually felt the need to revise their initial conclusions, in fact, they were as strong or stronger than what could be seen in the first iteration for companies C3 and C7 while no correlation was found for the assessments made by the representatives of company C1 who had attested to being spot on with their assessments based on the chat interviews.

Despite interviewers reporting instances of having arrived at inaccurate conclusions based on the chat, the correlations between the system ranks and the manually assigned ranks were strong for all companies, whereas the effect had been strong for only half of the evaluated companies in the first iteration with effect sizes on average also being higher in the second iteration.

While there is no clear explanation for the stronger correlations found examining the position tables of the second event, the reduction of questions in the chat interview from 6 to 4 might have played a role in it. Company reps had reported having been unable to

perform further inquiry in the first iteration due to the insufficient amount of time they had for each question (cf. Section 6.3.3). In the second iteration, they appeared not to have faced this problem. In fact, the time budget allotted to the four questions of the chat interview seemed to be perfectly adequate (cf. Section 7.2.1). Because of this, it stands to reason that company reps managed to form a more rounded impression of the candidates due to being able to engage with them more. A second factor that might have played into the system position table being more in line with the manual one, might be the fact that two out of the three participating companies had already attended the first installment of the event. Consequently, it seems conceivable that the learnings from their first participation helped them conduct the interviews more effectively.

Part 4

Analysis

CHAPTER 8

Discussion

The current chapter explores the key findings of the research project, analyzing them in the context of the theory outlined in Part 1 of this thesis. In addition to drawing from the theoretic background gathered during the initial literature research, an additional body of research was introduced in response to the subject matter of interview anxiety. This topic was an unanticipated finding that emerged from the thematic analysis of the applicant impulse interviews.

When trying to design a system for fairness and eliminate bias in the job interview process, the literature research provided a clear direction. Fairness in hiring hinges on two key principles: anonymity and comparability. Ideally, anonymity (cf. blind auditions in Section 3.3) and structured, horizontally scored (evaluating each question across all candidates before moving to the next) interviews with questions weighted according to their importance would create a system with high comparability and reduced cognitive bias. There were a few factors, however, that made it necessary to deviate from this gold standard approach. Hence, to address research question RQ1 these best practices had to be aligned with the domain-specific constraints and the needs of company representatives to arrive at a design that would ride the thin line between absolute fairness and sufficient stakeholder buy-in.

Limitations owed to the event format – a live recruiting event where multiple companies would interview a large group of students – additionally necessitated a few adaptations. The fact that multiple interviewers per company would interview students in parallel paired with the sheer number of interviewees didn't allow for the implementation of horizontal scoring. A second adjustment that was presupposed by the limitations in interview time allotted by the event format, was that the structured text interviews were not entirely spontaneous but based on the applicants' previously submitted answers. This mitigation emerged from the design workshop to address the expected problem of a text-based approach being too time-intensive to establish a meaningful exchange based on impromptu conversation alone. Finally, pre-study insights revealed companies'

deep skepticism towards anonymization with their current practices relying heavily on semi-structured or even unstructured interviews (cf. Section 4.2), online application tools with no built-in anonymization features, and diversity training as the sole measure to combat bias. This suggested that a strict blind hiring approach would be poorly received. To achieve greater acceptance, a two-tiered solution was devised: a fully anonymized, structured text interview followed by a traditional face-to-face interview. This should allow applicants a fair initial evaluation free from (most) unconscious biases. With these adaptations to the devised ideal process, I intended to strike a delicate balance between the mitigation of biases and stakeholder acceptance by focussing on three core concepts that are explained in detail in Section 5.3: *Structure*, *Comparability*, and *Anonymity*.

Looking at the evaluation data, the system showed great promise regarding the buy-in from company representatives, albeit not in every single case. While the majority of company reps found the chat interviews to work as intended, the user evaluations highlighted time constraints as a potential barrier to adoption, with some voices finding textual interviews to be too time-consuming (cf. Section 7.2.1). This uttered sentiment is reflective of the general skepticism towards the usage of anonymization in the hiring process that had presented itself in the pre-study, as mentioned in Section 4.2.3. After the adaptations in questionnaire scope made for the second iteration, only a single company mentioned issues with the time budget. While this might point towards a potential problem with the scope of said company's questionnaire, research shows that communication channels with a limited ability to convey nonverbal cues generally lead to a slower rate of exchanged information (cf. Section 3.3.1). Given this fact, it is important to manage expectations on what can be achieved in the given time frame, to allow company representatives to construct their questionnaires more effectively. When proposed with the alternatives of either increasing the time budget or reducing the number of questions in the questionnaire, company reps leaned toward the latter option.

When it came to fulfilling the company reps' selection needs while maintaining the applicants' anonymity in practice – the subject of research question RQ2 – it was necessary to examine the problem from two different angles: I had to evaluate if the company reps discovered what they were trying to learn in the interviews while also examining potential shortcomings to the anonymization efforts.

While an applicant's perceived skills certainly are an important factor in a hiring decision and competency-based arguments dominate their rationalization, they are by far not the only ones. In practice, employer decision-making is influenced by many considerations that are especially difficult to isolate, as a 2020 meta-study found [130]. When it came to the text-based employment interviews conducted in the course of this research, they were typically personality- rather than skill-focused.

In regards to meeting the hiring party's information needs, interviewers reported having been able to assess an applicant's personality through the chat, using factors like response times and typing behavior as cues to supplement the textual content. They claimed to form mental images of the applicants based on these textual interactions, often surprised by how these images clashed with the person they got to meet during the second interview

stage. While this surprise could indicate limitations of the chat format in fully capturing an applicant's personality, it could also be interpreted as evidence that the anonymization process functioned as intended by disrupting preconceived notions about applicants based on visual appearance. Some interviewers attributed these discrepancies to filling information gaps about the applicants with assumptions, essentially creating a potentially flawed mental image. This process was dubbed *schema activation* in 1975 by Marvin Minsky [131]. It describes how, whenever we encounter a new situation or stimulus, our brains activate existing knowledge structures, or schemas, that relate to that situation. These schemas provide a framework for interpreting the new information and filling in any gaps. For instance, some interviewers admitted to assuming their vis-a-vis male by default, potentially due to a prevalent stereotype: The default masculinity present in STEM fields [132]. Science is associated with a male person and masculine traits starting at an early age [133] – a stereotype that can translate to what company reps reported to have experienced: They pictured and talked about students as male by default only to be proven wrong upon meeting them when an applicant was not (c.f. Section 6.3.3):

> *"It wasn't so much that we were surprised. But implicitly you always talk about him. Him. That's more or less how it goes in technology."* [1]

The impact of such assumptions remains unclear – could a female student initially perceived as male be disadvantaged during the face-to-face interview, or even benefit from a lower initial bar? Further research could explore how these preconceptions influence later interactions. The possibility of the text-based interview challenging interviewer biases is an intriguing avenue for further research. Future studies could explore how different communication styles and personality traits manifest in text-based interactions, and how these influence interviewer perceptions.

With respect to maintaining the applicants' anonymity during the chat interview, the system performed well the majority of the time. However, the two-tiered design meant that inherent anonymity only extended to the first interview stage, leaving applicants potentially vulnerable to bias during the face-to-face interview. While this was a known limitation introduced as a concession to foster acceptance by the company reps, the user evaluation unearthed some further challenges in fully preventing attempts to undermine anonymity.

At times, interviewers made educated guesses about an applicant's gender based on word choice, as reported by a student in Section 6.3.3 who had her gender correctly inferred by the interviewers. While this might as well be attributed to pure chance, research suggests that it is possible to infer a person's gender from their writing with a certain level of reliability. The most successful method to predict author gender currently reaches an accuracy of 93,4%, according to a 2019 literature survey [134].

[1] Translated from German by the author. Original quote by a representative of company C1: "Es war jetzt nicht, dass man überrascht war. Aber man redet halt implizit immer von *ihm. Ihm.* In der Technik ist es halt ein bisschen so."

Other incidents included a direct question about an applicant's origin, interviewers introducing themselves with their names leading to applicant reciprocation, and one applicant giving up their identity on purpose when they were asked to describe themselves. These breaches highlight areas for improvement in future iterations. While these shortcomings could conceivably be addressed through technological means, providing the participants with clear instructions on what information not to disclose while setting a few ground rules for interviewers about what questions they can't ask would constitute a low-cost mitigation. While not guaranteeing anonymity, these measures should further reduce the likelihood of anonymity being broken. Despite the above-mentioned shortcomings, the prevalent assessment of the DEBIAS process' among applicants was that it successfully mitigated potential biases. See Section 7.2.1 for more detail on the applicant sentiments concerning this subject.

Research question RQ3 was tackled by statistically examining correlations between the companies' custom applicant rankings and the numerically determined rankings based on the assigned score values.

Encouragingly, the anonymized interview system, combining a structured text chat with a traditional face-to-face interview, demonstrated strong potential in predicting hiring decisions. In the second iteration, the combined system's ranking even outperformed the in-person interview for all companies (cf. Table 7.1), exhibiting a stronger correlation with the manually assigned rankings compared to the face-to-face interview alone. This suggests that, although company reps had the chance to interview all applicants in person, they still relied on what they had learned in the text interview as well to arrive at a more complete assessment.

For one company, even the correlation between the chat and manual rankings was slightly stronger than the one between the face-to-face interview and manual rankings. While the text chat alone seemed a weaker predictor, it still exhibited a moderate effect size for two out of three companies in the second iteration.

These findings are particularly noteworthy considering the initial limitations observed in the first iteration. Indeed, the correlations between manual and system ranks were moderate for two and strong for the two remaining companies but the face-to-face interview alone demonstrated a stronger correlation with the final ranking than the combination of chat and in-person interviews.

While the exact reasons for the stronger correlations in the second iteration remain unclear, several factors might be at play. The reduction in chat interview questions from six to four likely allowed company representatives to delve deeper into each question during the allotted time. This assumption is supported by the fact that time being short was not brought up by company reps in the interviews conducted during the second iteration and the explicit mention of the positive effects of reducing the number of questions per interview by one company rep who had participated in both events.

The increased time budget per question could in turn have facilitated a more comprehensive understanding of the applicants. Additionally, the experience gained from

participating in the first iteration might have contributed to more effective interview practices by the company representatives who returned for the second event.

Said values, however, still need to be taken with a grain of salt. Since there is no clear mechanism explaining how company reps arrived at their scores or their rankings, it is important to keep in mind that the comparability between companies and even between different raters from the same company is limited. In some cases, applicants who received great scores in the chat interview as well as the in-person interview were placed in the bottom half of the candidate field for the final ranking (cf. Section 6.3.4 and Section 7.2.2). This phenomenon cannot be explained by the company reps putting greater emphasis on personal interviews for their evaluations. If that was the case one would expect the rankings to be aligned strongly with the face-to-face scores. Instead, it might come down to individual differences in the rating schemes of company reps or the group dynamics between the teams of interviewers. What is considered a good performance can vary from one interviewer to the next. While one person scores more strictly, the next one might give more lenient ratings on average (cf. Section 2.2.8). All interviews were conducted in teams of two but all representatives of a company present at the event had to come together and arrive at a common position table for the final ranking. While this can help to even out leniency and severity biases, it is easy to imagine that some members might dominate the discussion and therefore manage to put the candidates they are most convinced of at the top. This is a factor that cannot be accounted for in the system design. How hard these inconsistencies are to interpret is exemplified by a case of a candidate receiving perfect scores for every question in the chat as well as the face-to-face interview with company C4 and hence receiving the top spot in the system ranking only to be ranked eleventh in the final manual position table.

Since Controlling for these influences was not possible due to the specifics of the recruiting event, the examination of the decision-making processes under the conditions of the socio-technical system presented in this work could provide an interesting avenue for further research.

In trying to answer research question RQ4, I was particularly surprised by how the text interviews were received by the applicants. The applicant evaluation yielded positive insights into the two-tiered interview process: The majority of students expressed confidence in their ability to present themselves effectively during the text chat interviews. This is particularly noteworthy given the potential challenges of self-presentation in a text-based format outlined in Section 3.3.1. Furthermore, many participants reported that the anonymized chat stage significantly reduced interview anxiety. This suggests that the ability to establish rapport with the interviewer while maintaining anonymity contributed to a more relaxed and comfortable interview experience.

The evaluation of the second iteration mirrored these findings. Students consistently reported feeling less nervous and more successful in presenting themselves due to the anonymized text chat. While not everybody seemed to appreciate the chat interview the same way, most candidates not only felt confident in their abilities to represent themselves in the chat interviews but even seemed to profit from the two-tiered approach

which showed potential in reducing nervousness. For them, the chat helped to establish a certain level of familiarity which, in turn, seemed to reduce the stress caused by being interviewed by a stranger. While the system ideally wouldn't impair a candidate's ability for self-representation, I had not expected it to possibly even improve upon it.

For many people, a job interview can be a very stressful situation. A set of inherent characteristics contribute to applicants experiencing feelings of anxiety in selection interview situations. The power imbalance in a typical selection interview, caused by the interviewer usually being in a position to grant something to the interviewee [135], can put applicants under a lot of pressure, especially, if they are in a position, where they are economically dependent on a positive outcome. Given that a job application is a highly evaluative process in general, especially the employment interview can evoke heightened feelings of anxiety in interviewees, due to it usually being largely out of their control. The interviewer being a stranger serves as an additional contributing factor, since, for individuals experiencing social anxiety, talking to strangers is especially difficult. [136].

Although social anxiety was not a topic explicitly covered during the impulse interviews, research suggests that the two-tier interview process characteristic of the DEBIAS design might especially benefit people experiencing symptoms of social anxiety – a portion of the population that has been growing for years, especially among the young [137]. The COVID-19 pandemic seems to have accelerated this trend even with low-income workers being particularly affected [138]. Hence, an increasing number of people might benefit from the anonymous computer-mediated interview process to alleviate their social anxiety.

A 2015 study comparing the symptoms of social anxiety in computer-mediated vs. face-to-face communication, found that, while computer-mediated communication did not actually alleviate physiological signs of anxiety when compared to in-person communication, individuals experiencing social anxiety still felt more comfortable during computer-mediated communication [139]. The study found that computer-mediated communication was perceived as more controllable and less threatening by participators and elicited a greater sense of success. Comforting features, such as relative anonymity and the absence of non-verbal features have been found to mitigate feelings of anxiety in initial interactions [77].

This apparent reduction of experienced stress might even lead to better outcomes for the candidates, since signs of stress, exposed during a selection interview, have been found to negatively impact applicant success, due to interviewers evaluating a candidate's competency lower, if they expose visible stress [140].

Other than the additional time to contemplate, an absence of nonverbal cues might have contributed to the quality of the answers because they did not need to monitor nonverbal behavior and hence were able to allocate more resources to the content of their messages (cf. Section 3.3.1).

While the lack of nonverbal features offers potential benefits in moderating social anxiety, it seems natural, that cutting out nonverbal aspects of communication comes with its

own set of challenges. Mandal summarizes the functions of nonverbal behavior in human communication as follows:

> *"Nonverbal signs help regulate the system, cueing hierarchy and priority among communicators, signaling the flow of interaction, and providing meta-communication and feedback."*[71]

Considering the above-mentioned role nonverbal play in communication, it is not entirely surprising that some communication challenges arose due to the text-based format. Applicants reported difficulty interpreting interviewers' intent without nonverbal cues, leading to occasional misunderstandings or insecurities about what the company reps thought about their answers. This could impact an applicant's interview performance.

One important predictor of someone's interview (as well as job) performance is their *Ability To Identify Criteria*, or *ATIC* for short [141]. This refers to their ability to identify the relevant performance criteria in situations where they are evaluated and said information is not made transparent. the evaluator's reactions are one crucial piece of information used to infer said criteria [142] which might suffer in a text-based interviewer environment and hence negatively affect the applicants' performances.

Some interviewers mitigated this by providing explicit verbal feedback in the chat, while others used emojis to enhance emotional context, both of which were very welcome by applicants. In general, the usage of emojis was very sparse. Whenever company reps using them came up in the interviews, though, students would mention them as being helpful for gauging interviewer sentiment, a function of emojis found in literature as well (cf. Section 3.3.1).

Additionally, the use of emojis seemed to create a less formal, more welcoming atmosphere. Support for this sentiment can be found in literature, according to which written communication is generally "[...] likely to be described as less friendly, emotional, or personal and more serious, businesslike, or task-oriented" [143]. Concerning emoticons, research suggests, that the usage of emojis leads to the message sender being perceived as friendlier and more affectionate [144].

The fact that the interviewers did apparently not notice miscommunication as a problem in the process might be attributed to their role in the hierarchy represented by the selection interview. Since the interviewers are in a position of power, effectively deciding if they are going to offer the interviewees a job [135], they have to be less concerned with how their messages are received. For them, there are far fewer, less momentous consequences in case their messages are misinterpreted.

The mentioned miscommunications provide further opportunities for future research. Fostering meta-communication and the usage of emoticons to supplement the textual content might help to prevent miscommunication and create a more pleasant atmosphere that helps to reduce applicant anxiety even further.

These findings offer valuable insights into the potential of anonymized interview systems for promoting fairness and reducing bias in the hiring process. The system's demonstrated ability to generate rankings closely aligned with those produced by traditional interview methods, coupled with the potential for mitigating bias, suggests a promising path forward. Further research is warranted to explore the long-term potential of such systems in scenarios that more closely resemble the hiring processes used to fill a specific open position.

8.1 Limitations

In the following, I will summarize the above-mentioned limitations of this research, supplementing them with information on potential additional limitations and opportunities to address them in future research.

First, the recruiting event serving as the basis for the user evaluation itself is not necessarily representative of a typical hiring process. Company reps did not receive CVs or motivation letters and all participants were students. Additionally, there were no job descriptions provided to the students, meaning they did not interview for a specific position. Furthermore, the fact that the event was specifically targeted at students from underrepresented groups led to the relative proportion of those groups being much higher than what you would have typically found in a recruiting event, making it difficult to draw comparisons.

Second, this research did not examine the amount of discrimination present in face of the DEBIAS system or compare it to levels of discrimination in comparable recruitment formats. For lack of data, I could not establish a baseline measure of discrimination or examine the discrimination found in other events the *TU Career Center* had alleged. I had to rely on anecdotal evidence from the *TU Career Center* suggesting a history of discrimination, but a more objective measure might have strengthened the research.

Third, the impact of interviewers' surprise upon meeting applicants with contrasting physical characteristics remains unclear. Future research could explore the long-term effects of such surprises on hiring decisions.

Finally, potential rater biases such as leniency or severity could not be eliminated entirely. With each company conducting two interviews in parallel, an applicant's ranking might be influenced by the specific interviewer assigned, as some might be more lenient in their scoring than others. Additionally, the final ranking process within each company involved group discussions, potentially introducing group dynamics and cognitive biases. Ideally, these factors would be controlled for to lend additional validity to the results.

These limitations highlight the need for further research in controlled settings that more closely resemble real-world hiring practices. Future studies should incorporate more representative applicant pools, established baselines for discrimination, and standardized scoring mechanisms to strengthen the generalizability of the findings.

List of Figures

List of Tables

Appendix A

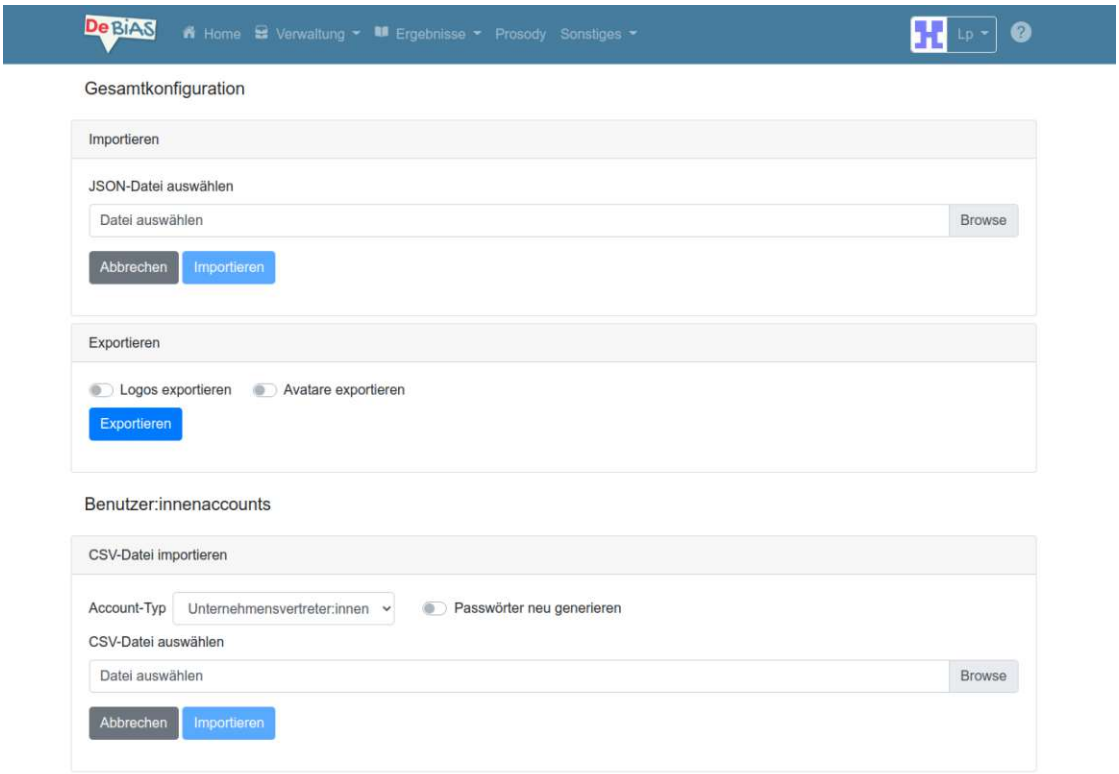


Figure 1: Import and export of configuration data such as user accounts

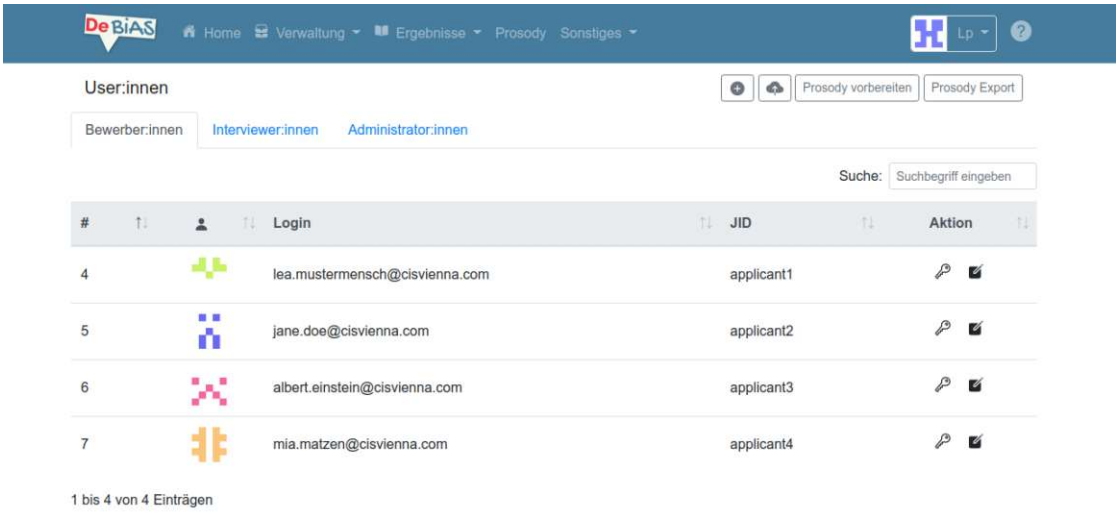


Figure 2: User list providing an overview of all user accounts by account type

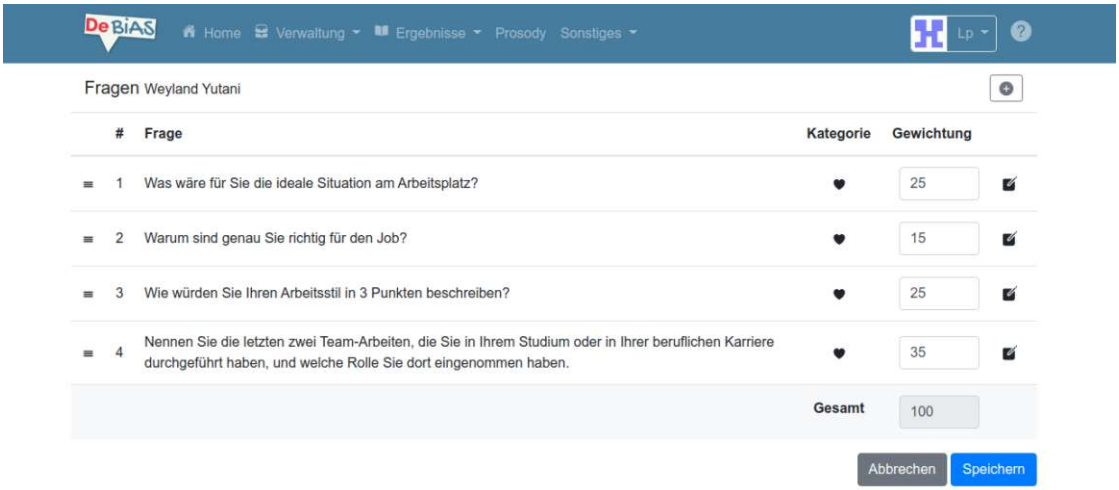


Figure 3: User interface for editing a questionnaire

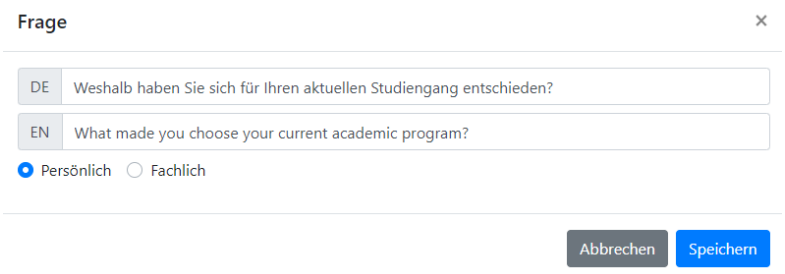


Figure 4: Modal dialog for editing a single question in a questionnaire

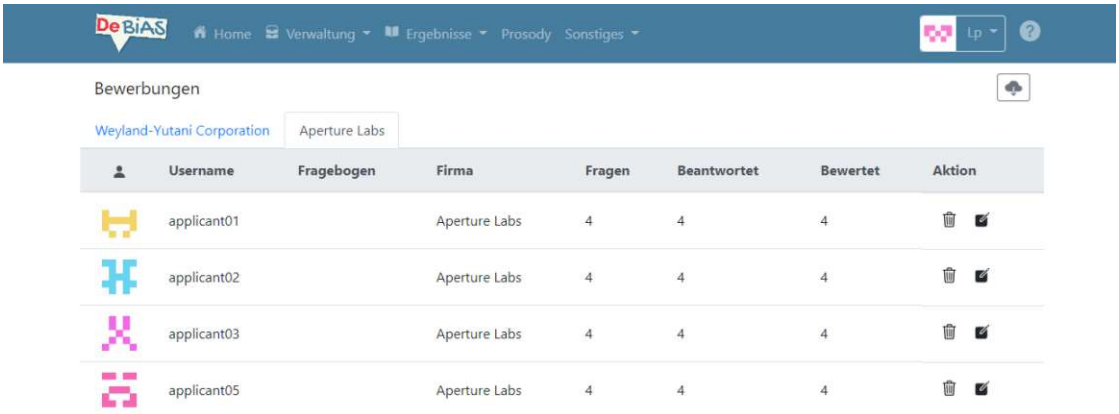


Figure 5: Tabluar overview of applicant ratings grouped by company

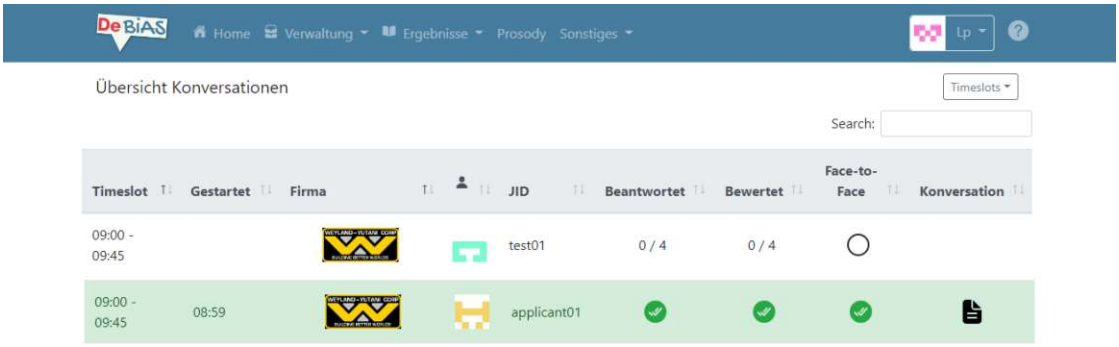


Figure 6: Overview of the process statuses for each applicant-company pairing

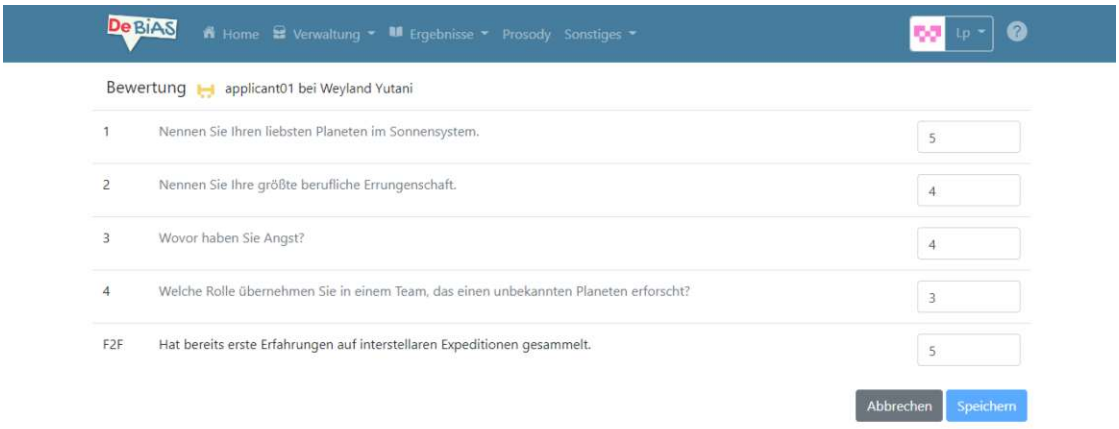


Figure 7: The tool allows administrators to edit applicant ratings at the request of company reps to correct input errors

Appendix B

Interviewleitfaden Vorstudie

Forschungsfrage

Was ist die Motivation für Firmen, sich des Themas „Unconscious Bias“ anzunehmen?

Einstieg

- Begrüßung, Dank für die Zeit
- Erklärung der Zielsetzung des Projekts
- Ablauf des Interviews (Dauer)
- Datenschutz

Fragen

Einstieg

Welche Beziehung zum TU Career Center und welche Erfahrungen damit gemacht.

Frage 1

Wie lange beschäftigen Sie sich bereits mit dem Thema Diskriminierung/Bias im Recruitmentprozess?

Wodurch sind Sie auf das das Thema selbst erstmals aufmerksam geworden?

Frage 2

Beschreiben Sie bitte den typischen Prozess, den ein*e Bewerber*in durchlaufen würde, um sich für eine Stelle in Ihrer Firma zu bewerben.

Frage 3

Welche Maßnahmen sind in dem Prozess implementiert, um eventuelle Biases zu berücksichtigen?

Rückfragen

Wenn keine besonderen Maßnahmen

Wie sehen die Pläne aus, das Thema zu adressieren?

Wenn Maßnahmen:

> Wie haben die getroffenen Maßnahmen die Auswahl von Bewerber*innen beeinflusst?

Wenn keine Maßnahmen:

> Welche Maßnahmen würden Sie ergreifen, wenn Sie den idealen Prozess zur Bekämpfung von Biases entwerfen müssten?

Frage 4

Worin liegen Ihrer Meinung nach die Probleme des klassischen Bewerbungsprozesses? (Bewerbung mit Foto, persönlichen Daten)

Frage 5

Was bedeutet für Sie „Diversity“ und welche Vorteile können sich dadurch ergeben?

Frage 6

Was versprechen Sie sich von Unternehmensseite davon, etwas gegen (eventuelle) unbewusste Biases zu unternehmen?

Frage 7

Was sind die Kriterien anhand derer in Vorstellungsgesprächen bewertet wird?

Rückfragen

Wie bekommen Sie diese Information?

Was wird gefragt, worauf wird geachtet?

Frage 8

Welche persönlichen Komponenten spielen eine Rolle, bei der Bewertung von Bewerber*inne*n?

Rückfrage

Welche Komponenten sind wie wichtig?

Frage 9

Was passiert, wenn man diese Komponenten ausblendet?

Frage 10

Wenn Sie das Ideale System zur Bekämpfung unbewusster Biases entwerfen müssten, wie würde das aussehen?

Frage 11

Werden in Ihrem Unternehmen anonymisierte Bewerbungsverfahren genutzt?

Rückfragen

Wenn nein:

Wie ist Ihre Einschätzung darüber?

Wie würden Sie das machen?

Wenn ja:

Wie wurde das eingeführt?

Wie ist die konkrete Umsetzung?

Welche Erfahrungen wurden damit gemacht?

Wie schätzen Sie die Situation in Österreich ein – sind anonymisierte Verfahren eher Regel oder Ausnahme?

Welche Gründe gibt es dafür, dass das nicht flächendeckend Verwendung findet?

Abschluss

Dank für Zeit, Zusammenfassung, Ausblick

Appendix C

Impuls-Interview Leitfaden Firmen

1. Wie ist der Chat Ihrer Meinung nach gelaufen?
   How do you think the chat worked out?

2. Konnten Sie sich während des Chats ein gutes Bild von den BewerberInnen machen?
   Could you assess the applicants well during the chat?

3. Gab es Beispiele, wo ihre Bewertung zwischen Chat und persönlichem Gespräch signifikant unterschiedlich war, und wenn ja, warum?
   Were there any examples where your assessment differed significantly between the chat and personal interview, and if so, why?

4. Gab es Dinge, die Ihnen beim Verwenden des Systems gefehlt haben (Features, Informationen, Tools)?
   Did you miss anything when using the system (features, informations, tools)?

5. Wollen Sie uns sonst noch etwas zu Ihren Erfahrungen heute sagen?
   Would you like to share anything else about your experience?

Appendix D

Impuls-Interview Leitfaden Studierende

1. Wie ist der Chat Ihrer Meinung nach gelaufen?
   How do you think the chat worked out for you?

2. Hatten Sie das Gefühl, sich durch die Fragenbeantwortung und im Chat gut präsentieren zu können?
   Do you feel you managed to represent yourself well by answering the questions and during the chat?

3. Was waren Ihrer Meinung nach die größten Unterschiede zwischen dem Chat und dem persönlichen Gespräch?
   What did you perceive to be the biggest differences between the Chat and the personal interview?

4. Gab es Dinge, die Ihnen beim Verwenden des Systems gefehlt haben (Features, Informationen, Tools)?
   Did you miss anything when using the system (features, informations, tools)?

5. Wollen Sie uns sonst noch etwas zu Ihren Erfahrungen heute sagen?
   Would you like to share anything else about your experience?

Appendix E

Impuls-Interview Leitfaden Studierende

1. Wie ist der Chat deiner Meinung nach gelaufen?
   How would you say the chat went for you?

2. Hattest du das Gefühl dich im Chat selbst gut repräsentieren zu können?
   Do you feel like you were able to represent yourself well during the chat?

3. Hast du bereits anderweitig Erfahrung mit Bewerbungsgesprächen?
   Have you previously had experiences with job interviews?

   a. Was war für dich der größte Unterschied zu deinen bisherigen Erfahrungen?
      What was the biggest difference between the process today and your previous experiences?
   b. Hast du den Eindruck, dass es möglich ist mit diesem Prozess den Einfluss von Biases zu verringern?
      Do you think the process employed today can actually help lessen the impact of biases?

4. Gibt es etwas, das du an dem System gerne ändern würdest oder etwas, das dir gefehlt hat?
   Did you think that the tool that you used was missing any features? Is there something you would have liked to have seen?

5. Willst du noch etwas anderes loswerden zu der Veranstaltung, dem Ablauf oder sonstigem?
   Any other impression you want to share?

Appendix F

Einverständniserklärung

Die/Die Unterzeichnende erklärt sich bereit, am Forschungsprojekt „**DEBIAS - Digitally Eliminating Bias in Applicant Selection“** im Rahmen des **Voice of Diversity Events** am **07.10.2020** teilzunehmen.

Die/Der Unterzeichnende nimmt zur Kenntnis, dass im Zuge der Studie wissenschaftliche Daten in Form von Interviews, Video- und Audio-Aufzeichnungen, sowie Notizen gesammelt werden. Weiters wird für das Event das prototypische „DEBIAS-Tool“ eingesetzt, welches die folgenden Daten speichert:

- Name, Email-Adresse
- Fragenbeantwortungen & Bewertungen
- Chatverlauf
- Metadaten: Zuordnung TeilnehmerIn - Firma

Diese werden **streng vertraulich** behandelt und **keinesfalls mit Dritten geteilt**. Im Falle einer etwaigen wissenschaftlichen Publikation werden die erhobenen Daten jedenfalls nur in aggregierter Form oder anonymisiert publiziert.

Die/Der Unterzeichnende kann dieser Einverständniserklärung jederzeit widerrufen und eine Löschung aller bis dahin gesammelten Daten erwirken. Dies führt jedoch zum Ausscheiden aus der Studie, da die Sammlung und Auswertung dieser Daten die Voraussetzung für die wissenschaftliche Durchführung des Projekts darstellt.

Wien, 07.10.2020

_____

[Name Unterzeichnende*r]

_____

[Unterschrift]

Declaration of consent

The undersigned agrees to participate in the research project "**DEBIAS - Digitally Eliminating Bias in Applicant Selection**" as part of the Voice of Diversity Event on 07.10.2020.

The undersigned acknowledges that in the course of the study scientific data will be collected in the form of interviews, video and audio recordings, and notes. Furthermore, the prototypical "DEBIAS-Tool" will be used which stores the following data:

- Name, email address
- Question answers & ratings
- Chat history
- Metadata: Assignment participant - company

These will be kept **strictly confidential** and **will not be shared with third parties under any circumstances**. In the event of any scientific publication, the collected data will only be published in aggregated or anonymized form.

The undersigned may revoke this declaration of consent at any time and obtain the deletion of all data collected up to that point. However, this will lead to the exclusion from the study, as the collection and analysis of these data is a precondition for the scientific implementation of the project.

Vienna, October 7th, 2020

_____

[Name of the undersigned]

_____

[Signature]

Appendix G

Einverständniserklärung

Die/Die Unterzeichnende erklärt sich bereit, am Forschungsprojekt „**DEBIAS - Digitally Eliminating Bias in Applicant Selection“** im Rahmen des **Voice of Diversity Events** am **08.06.2021** teilzunehmen.

Die/Der Unterzeichnende nimmt zur Kenntnis, dass im Zuge der Studie wissenschaftliche Daten in Form von Interviews, Video- und Audio-Aufzeichnungen, sowie Notizen gesammelt werden. Weiters wird für das Event das prototypische „DEBIAS-Tool“ eingesetzt, welches die folgenden Daten speichert:

- Name, Email-Adresse
- Fragenbeantwortungen & Bewertungen
- Chatverlauf
- Metadaten: Zuordnung TeilnehmerIn - Firma

Diese werden **streng vertraulich** behandelt und **keinesfalls mit Dritten geteilt**. Im Falle einer etwaigen wissenschaftlichen Publikation werden die erhobenen Daten jedenfalls nur in aggregierter Form oder anonymisiert publiziert.

Die/Der Unterzeichnende kann dieser Einverständniserklärung jederzeit widerrufen und eine Löschung aller bis dahin gesammelten Daten erwirken. Dies führt jedoch zum Ausscheiden aus der Studie, da die Sammlung und Auswertung dieser Daten die Voraussetzung für die wissenschaftliche Durchführung des Projekts darstellt.

Wien, 08.06.2021

_____

[Name Unterzeichnende*r]

_____

[Unterschrift]

Declaration of consent

The undersigned agrees to participate in the research project **"DEBIAS - Digitally Eliminating Bias in Applicant Selection"** as part of the Voice of Diversity Event on 08.06.2021.

The undersigned acknowledges that in the course of the study scientific data will be collected in the form of interviews, video and audio recordings, and notes. Furthermore, the prototypical "DEBIAS-Tool" will be used which stores the following data:

- Name, email address
- Question answers & ratings
- Chat history
- Metadata: Assignment participant - company

These will be kept **strictly confidential** and **will not be shared with third parties under any circumstances**. In the event of any scientific publication, the collected data will only be published in aggregated or anonymized form.

The undersigned may revoke this declaration of consent at any time and obtain the deletion of all data collected up to that point. However, this will lead to the exclusion from the study, as the collection and analysis of these data is a precondition for the scientific implementation of the project.

Vienna, June 8th, 2021

_____

[Name of the undersigned]

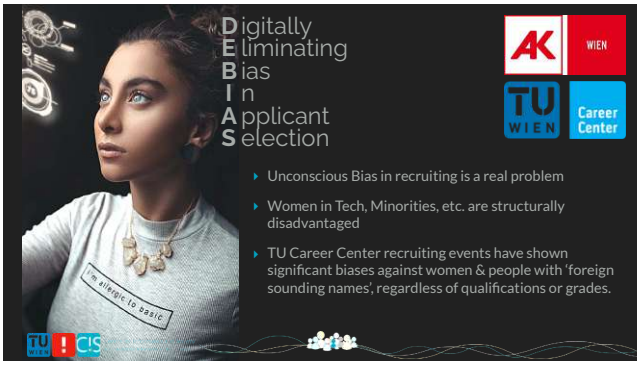
_____

[Signature]

Appendix H

Brainstorming Socio-Technical Solutions

DEBIAS



Some *more* facts about Bias...

- Unstructured Interviews are „... *the greatest failure of industrial and organizational psychology*"
  - It's really hard to get employers to rely on technological decision aids...

Some *more* facts about Bias...

- Unstructured Interviews are „... *the greatest failure of industrial and organizational psychology*"
  - It's really hard to get employers to rely on technological decision aids...
    ... even though there is an overwhelming amount of data showing that they **simply do not work** when trying to combat bias!

Some *more* facts about Bias...

- Unstructured Interviews are „... *the greatest failure of industrial and organizational psychology*"
  - It's really hard to get employers to rely on technological decision aids...
    ... even though there is an overwhelming amount of data showing that they **simply do not work** when trying to combat bias!
- *Confirmation Bias*
- *Peak-End Rule* or *Recency Bias*
  - most intense or recent experiences are weighed more strongly in decisions
- *Groupthink*
  - Shared evaluations can be biased because they tend towards uniformity

„What Works": Some Strategies To Combat Bias

- Remove identifying and demographic information, photos, names, etc. from applications and the recruiting process
- Enforce structured interviews
  - Predetermine questions, grading, and weight of sought after qualities
  - Evaluate immediately after each question / topic
  - If multiple evaluators are involved, ensure they evaluate separately first
- Allow comparative evaluations rather than sequential ones
  - Evaluate answers side-by-side and comparatively

BRAINSTORMING SOLUTIONS

- Empirical methodology
  - Qualitative Interviews with 5 companies partaking in the event
  - Survey of students towards their experiences w/ bias and discrimination in the recruitment process
  - 2 Workshops w/ recruiters before the event
- Event organisation
  - 5 companies with multiple interviewers each, ~50 students
  - Location: Kuppelsaal
  - Whole-Day Event

- What social or technological solutions could be developed to ...
  ... combat conscious and unconscious biases on the side of the recruiters?
  ... protect applicants' identities and demographic data?
  ... allow the setting between applicant and recruiter to be as close as possible to a conventional recruiting talk and raise future acceptance of the techniques developed?
- How can the given solutions be evaluated at or after the event?

Appendix I

Digitally Eliminating Bias in Applicant Selection

DEBIAS Tool

Lukas Pichlhöfer
lukas.pichlhoefer@tuwien.ac.at

Florian Cech
florian.cech@tuwien.ac.at

Einführung

Agenda

- Warum DEBIAS?
- Verwendung des Tools
- Wissenschaftliche Begleitforschung

Warum DEBIAS?

DEBIAS...

- hilft, Teile des Recruitingprozesses zu **anonymisieren**
- trägt dazu bei, den Interviewprozess zu **strukturieren**
- unterstützt die **Entscheidungsfindung** durch vergleichbare Bewertung der Performance der BewerberInnen
- ist Grundsätzen der Fairness und Gleichbehandlung unterworfen

3 Standbeine

- reduziert Biases, die aufgrund persönlicher Eigenschaften (Aussehen, Gender, ethnischer Herkunft) auftreten
- erlaubt eine erstes Assessment persönlicher und fachlicher Kompetenzen
- stellt in DEBIAS eine Ergänzung zum klassischen, face to face Interview dar

Anonymisierung
Strukturierung
Entscheidungsfindung

3 Standbeine

- definiert klare, gleichbleibende Regeln und Abläufe für den Prozess
- erlaubt echte Vergleichbarkeit zwischen verschiedenen InterviewpartnerInnen herzustellen
- reduziert das Potential für weitere Biases (Anchoring, Clustering & Salience)

Anonymisierung
Strukturierung
Entscheidungsfindung

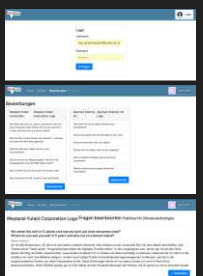
3 Standbeine

- erleichtert nicht nur sequenzielle, sondern paralelle Bewertung der Performance
- erlaubt Vergleich auf Basis einzelner Fragen / Themen und Antworten
- evaluiert die eigene Einschätzung auch im Vergleich mit dem persönlichen Gespräch

Anonymisierung
Strukturierung
Entscheidungsfindung

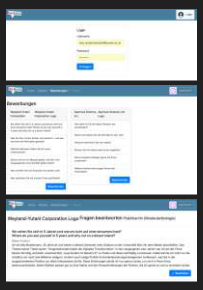
Eckdaten des Systems

- Prototyp - keine fertige, komplette Software
- Umsetzung als Webapplikation
- 2 Teilbereiche
  - Interviewer:innen
  - Bewerber:innen
- Personalisierte Zugangsdaten

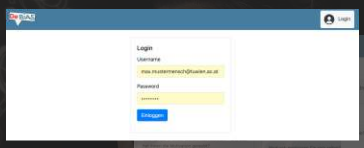


Ablauf des Bewerbungsprozesses

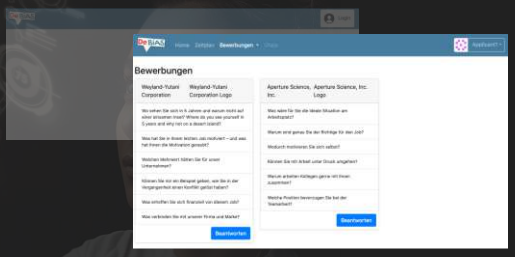
1. Bewerber:innen beantworten Fragen
2. Interviewer:innen initiieren digitales Bewerbungsgespräch (Chat)
3. Fragen und Chat-Antworten werden einzeln bewertet
4. Persönliches, face-to-face Gespräch ergänzt den Chat
5. Interviewer:innen vergleichen Performance der Bewerber:innen anhand von Bewertungsergebnissen



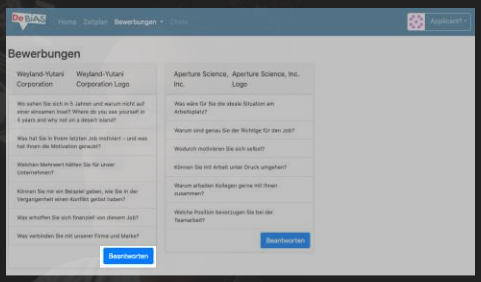
Schritt 1: Login & Übersicht über Bewerbungen



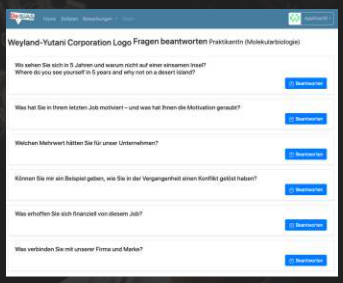
Schritt 1: Login & Übersicht über Bewerbungen



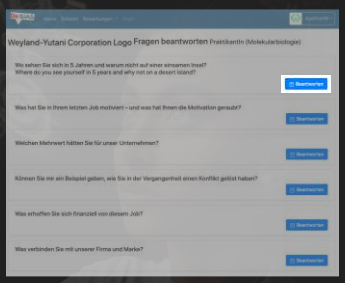
Schritt 2: Fragenbeantwortung



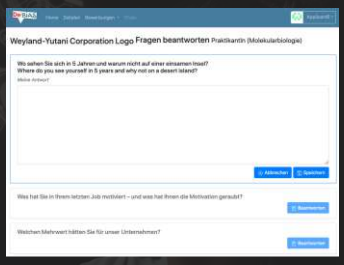
Schritt 2: Fragenbeantwortung



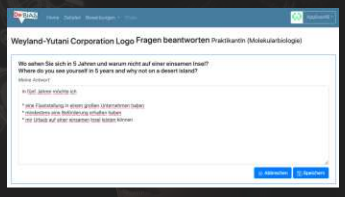
Schritt 2: Fragenbeantwortung



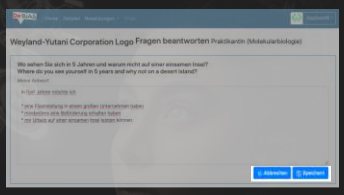
Schritt 2: Fragenbeantwortung



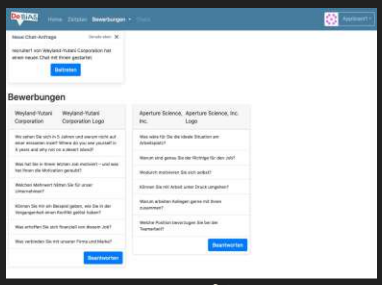
Schritt 2: Fragenbeantwortung



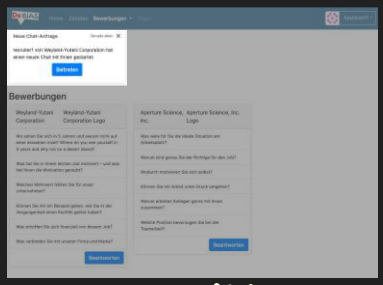
Schritt 2: Fragenbeantwortung



Schritt 2: Start des Chats

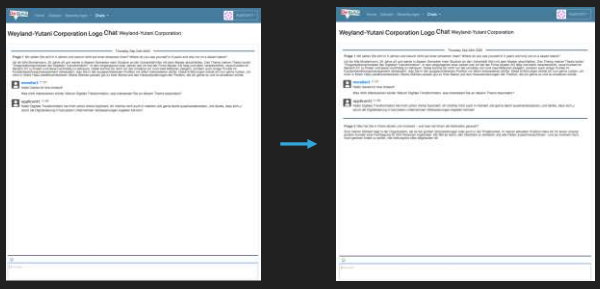


Schritt 2: Start des Chats

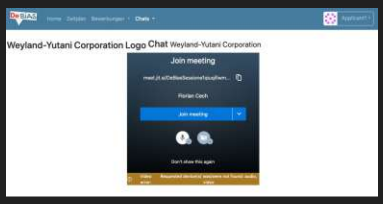


Schritt 2: Start des Chats

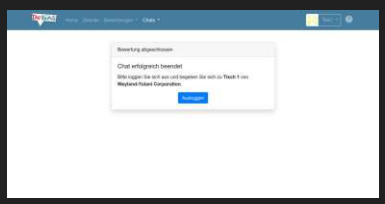
Schritt 2: Start des Chats



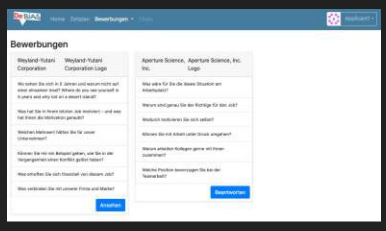
Schritt 3:
Persönliches Gespräch (online)



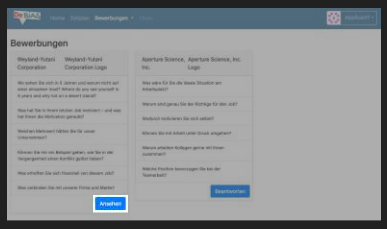
Schritt 3:
Persönliches Gespräch (vor Ort)



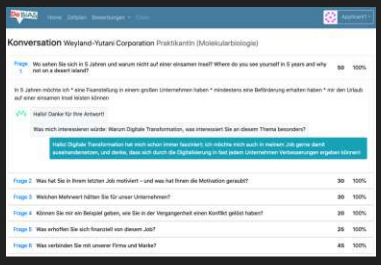
Schritt 4:
Gespräch nachlesen & Bewertungen einsehen



Schritt 4:
Gespräch nachlesen & Bewertungen einsehen



Schritt 4:
Gespräch nachlesen & Bewertungen einsehen



Wissenschaftliche Begleitung

- DEBIAS Tool wird wissenschaftlich im Rahmen des Voice of Diversity Events evaluiert
- Einverständnis zur Datensammlung und Auswertung
  - Selbstverständlich komplett anonymisiert
- Vorgehen beim Event:
  - Stichprobenartige Videobeobachtung
  - Kurzinterviews im Laufe des Tages
  - Begleitung des Entscheidungsprozesses

Wissenschaftliche Begleitung: Evaluierung

- Ziel der Studie ist Design und Evaluierung des DEBIAS Tools
  - Wo liegen Stärken und Schwächen des Systems?
  - Können Schwächen durch weiter technologische oder soziale Massnahmen ausgeglichen werden?
  - Lassen sich positive Effekte in Bezug auf Fairness bei der Auswahl beobachten?
  - Wie wird das System von Bewerber:innen und Interviewer:innen aufgenommen?

Besten Dank für die Aufmerksamkeit!

Fragen?

Projektwebsite:
https://debias.cisvienna.com

Bibliography

[1] S. Austria, "Erwerbstätigkeit." [Online]. Available: https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/soziales/gender-statistik/erwerbstaetigkeit/index.html [Accessed: 2021-09-10]

[2] ——, "Demographische Prognosen," Nov. 2019. [Online]. Available: http://www.stat.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/demographische_prognosen/122071.html [Accessed: 2021-09-13]

[3] T. Fent, A. Fürnkranz-Prskawetz, B. Hammer, and G. Danhel, "Demographischer Wandel - geänderte Rahmenbedingungen für den Sozialstaat?" Bundesministerium für Arbeit, Sozial, Gesundheit und Konsumentenschutz, Tech. Rep., 2019.

[4] WKO, "Demografische Entwicklung in Österreich," May 2020. [Online]. Available: https://news.wko.at/news/oesterreich/Demografische_Entwicklung_in_Oesterreich.html [Accessed: 2021-09-10]

[5] L. Lippens, S. Vermeiren, and S. Baert, "The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments," *European Economic Review*, vol. 151, p. 104315, Jan. 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0014292122001957 [Accessed: 2024-03-14]

[6] A. Reynolds and D. Lewis, "Teams Solve Problems Faster When They're More Cognitively Diverse," Mar. 2017. [Online]. Available: https://hbr.org/2017/03/teams-solve-problems-faster-when-theyre-more-cognitively-diverse [Accessed: 2024-03-14]

[7] V. Hunt, D. Layton, S. Prince, and others, "Diversity matters," *McKinsey & Company*, vol. 1, no. 1, pp. 15–29, 2015, publisher: New York, NY, USA:.

[8] J. Bersin, "Why Diversity and Inclusion Has Become a Business Priority," Dec. 2015. [Online]. Available: https://joshbersin.com/2015/12/why-diversity-and-inclusion-will-be-a-top-priority-for-2016/ [Accessed: 2024-03-14]

[9] C. Herring, "Does Diversity Pay?: Race, Gender, and the Business Case for Diversity," *American Sociological Review*, vol. 74, no. 2, pp. 208–224, Apr. 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/000312240907400203 [Accessed: 2024-03-14]

[10] K. Bezrukova, C. S. Spell, J. L. Perry, and K. A. Jehn, "A meta-analytical integration of over 40 years of research on diversity training evaluation." *Psychological Bulletin*, vol. 142, no. 11, pp. 1227–1274, Nov. 2016. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/bul0000067 [Accessed: 2023-01-03]

[11] C. Goldin and C. Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, vol. 90, no. 4, pp. 715–741, Sep. 2000. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257%2Faer.90.4.715&source=post_page-----deda93da109b---------- [Accessed: 2021-01-12]

[12] O. Åslund and O. N. Skans, "Do Anonymous Job Application Procedures Level the Playing Field?" *ILR Review*, vol. 65, no. 1, pp. 82–107, Jan. 2012, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/001979391206500105 [Accessed: 2021-01-10]

[13] M. Kreuzhuber and J. Moreno-Hasenöhrl, "Anonymisierte Bewerbungen: Defensiv-strategie mit Nachteilen für Unternehmen," May 2016. [Online]. Available: https://news.wko.at/news/oesterreich/Anonymisierte_Bewerbungen.html [Accessed: 2021-02-18]

[14] P. Mayring, "Qualitative Inhaltsanalyse," in *Handbuch qualitativer Forschung: Grundlagen, Konzepte, Methoden und Anwendungen*. Beltz Verlag, Weinheim und Basel, 1991, pp. 209–213.

[15] C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[16] C. D. Hardin and M. R. Banaji, "The nature of implicit prejudice," in *The behavioral foundations of public policy*. Princeton University Press, 2013, pp. 13–31.

[17] K. Posey, "Green Book (2018) and BlacKKKlansman (2018): An Analysis of White and Black Perspectives in Contemporary Films Using Critical Race Theory," *Honors College Theses*, Jan. 2020. [Online]. Available: https://digitalcommons.georgiasouthern.edu/honors-theses/507

[18] A. G. Greenwald and L. H. Krieger, "Implicit Bias: Scientific Foundations," *California Law Review*, vol. 94, no. 4, p. 945, Jul. 2006. [Online]. Available: https://www.jstor.org/stable/10.2307/20439056?origin=crossref [Accessed: 2021-08-09]

[19] P. G. Devine, "Stereotypes and prejudice: Their automatic and controlled components." *Journal of personality and social psychology*, vol. 56, no. 1, p. 5, 1989, publisher: American Psychological Association.

[20] A. S. Baron and M. R. Banaji, "The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood," *Psychological science*, vol. 17, no. 1, pp. 53–58, 2006, publisher: SAGE Publications Sage CA: Los Angeles, CA.

[21] A. G. Greenwald, D. E. McGhee, and J. L. K. Schwartz, "Measuring individual differences in implicit cognition: The implicit association test." *Journal of Personality and Social Psychology*, vol. 74, no. 6, pp. 1464–1480, 1998. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.74.6.1464 [Accessed: 2023-04-13]

[22] A. G. Greenwald and M. R. Banaji, "The implicit revolution: Reconceiving the relation between conscious and unconscious." *American Psychologist*, vol. 72, no. 9, pp. 861–871, Dec. 2017. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/amp0000238 [Accessed: 2021-08-09]

[23] D. Chugh, "Societal and managerial implications of implicit social cognition: Why milliseconds matter," *Social Justice Research*, pp. 203–222, 2004.

[24] M. Bursell, "What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process," Stockholm University, Faculty of Social Sciences, Department of Sociology, Stockholm, Working Paper 2007:7, 2007. [Online]. Available: http://su.diva-portal.org/smash/record.jsf?pid=diva2%3A821429&dswid=-2289

[25] M. Bertrand and S. Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, vol. 94, no. 4, pp. 991–1013, Sep. 2004. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/0002828042002561 [Accessed: 2021-01-12]

[26] E. Zschirnt and D. Ruedin, "Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015," *Journal of Ethnic and Migration Studies*, vol. 42, no. 7, pp. 1115–1134, May 2016. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/1369183X.2015.1133279 [Accessed: 2021-08-11]

[27] H. Hofer, G. Titelbach, D. Weichselbaumer, and R. Winter-Ebmer, "Diskriminierung von MigrantInnen am österreichischen Arbeitsmarkt," Institute for Advanced Studies, Vienna, Research report, 2013, publisher: Institut für höhere Studien.

[28] M. Piazolo, "Hiring Discrimination in the Labor Market," in *ICSB World Conference Proceedings*, Argentina, 2017, p. 10.

[29] R. E. Steinpreis, K. A. Anders, and D. Ritzke, "The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study," *Sex roles*, vol. 41, no. 7, pp. 509–528, 1999, publisher: Springer.

[30] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman, "Science faculty's subtle gender biases favor male students," *Proceedings of the National Academy of Sciences*, vol. 109, no. 41, pp. 16 474–16 479, Oct. 2012. [Online]. Available: http://www.pnas.org/cgi/doi/10.1073/pnas.1211286109 [Accessed: 2021-08-16]

[31] A. Booth and A. Leigh, "Do employers discriminate by gender? A field experiment in female-dominated occupations," *Economics Letters*, vol. 107, no. 2, pp. 236–238, May 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165176510000352 [Accessed: 2021-08-16]

[32] D. S. Hamermesh and J. E. Biddle, "Beauty and the Labor Market," *The American Economic Review*, vol. 84, no. 5, pp. 1174–1194, 1994, publisher: American Economic Association. [Online]. Available: http://www.jstor.org/stable/2117767 [Accessed: 2023-04-13]

[33] D. S. Hamermesh, *Beauty pays: why attractive people are more successful*. Princeton, NJ ; Oxford: Princeton University Press, 2011.

[34] J. Borland and A. Leigh, "Unpacking the Beauty Premium: What Channels Does It Operate Through, and Has It Changed Over Time?" *Economic Record*, vol. 90, no. 288, pp. 17–32, Mar. 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/1475-4932.12091 [Accessed: 2023-01-03]

[35] E. L. Thorndike, "A constant error in psychological ratings," *Journal of applied psychology*, vol. 4, no. 1, pp. 25–29, 1920, publisher: Whashington.

[36] M. G. Haselton and D. Buss, "Design Flaws or Design Features," *Social judgments: Implicit and explicit processes*, vol. 5, p. 23, 2003.

[37] H. A. Simon, "Bounded Rationality," in *Utility and Probability*, J. Eatwell, M. Milgate, and P. Newman, Eds. London: Palgrave Macmillan UK, 1990, pp. 15–18. [Online]. Available: https://doi.org/10.1007/978-1-349-20568-4_5

[38] A. Wilke and R. Mata, "Cognitive bias," in *Encyclopedia of human behavior*. Academic Press, 2012, pp. 531–535.

[39] A. Kramer and A. Harris, "Overcoming Gender And Affinity Biases in the Medical Profession," *An Evolution of Empowerment*, p. 34, 2020.

[40] R. Dobelli, *Die Kunst des klaren Denkens*. Hanser München, 2011.

[41] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases." *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974, place: US Publisher: American Assn for the Advancement of Science.

[42] R. Pohl, Ed., *Cognitive illusions: a handbook on fallacies and biases in thinking, judgement and memory*. Hove ; New York: Psychology Press, 2004, oCLC: ocm55124398.

[43] T. D. Wilson, C. E. Houston, K. M. Etling, and N. Brekke, "A new look at anchoring effects: basic anchoring and its antecedents." *Journal of Experimental Psychology: General*, vol. 125, no. 4, p. 387, 1996, publisher: American Psychological Association.

[44] D. Zhang and S. Highhouse, "Judgment and Decision Making in the Workplace," in *The SAGE handbook of industrial, work & organizational psychology: Personnel psychology and employee performance*, 2nd ed. Sage Publishing, 2018, vol. 1, pp. 611–633.

[45] G. Bierbrauer, *Sozialpsychologie*. W. Kohlhammer, 1996.

[46] E. E. Jones and V. A. Harris, "The attribution of attitudes," *Journal of Experimental Social Psychology*, vol. 3, no. 1, pp. 1–24, Jan. 1967. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0022103167900340 [Accessed: 2021-09-10]

[47] K. Jonas, W. Stroebe, M. Hewstone, and others, *Sozialpsychologie*. Springer, 2014.

[48] K. Dion, E. Berscheid, and E. Walster, "What is beautiful is good." *Journal of personality and social psychology*, vol. 24, no. 3, p. 285, 1972, publisher: American Psychological Association.

[49] D. Kahneman, *Thinking, fast and slow.* Macmillan, 2011.

[50] A. J. Schuh, "Contrast effect in the interview," *Bulletin of the Psychonomic Society*, vol. 11, no. 3, pp. 195–196, Mar. 1978. [Online]. Available: http://link.springer.com/10.3758/BF03336806 [Accessed: 2021-09-08]

[51] F. C. Lunenburg, "Performance appraisal: Methods and rating errors," *International journal of scholarly academic intellectual diversity*, vol. 14, no. 1, pp. 1–9, 2012.

[52] G. Huberman, "Familiarity Breeds Investment," *The Review of Financial Studies*, vol. 14, no. 3, pp. 659–680, Jul. 2001. [Online]. Available: https://doi.org/10.1093/rfs/14.3.659 [Accessed: 2021-09-10]

[53] A. Serenko and N. Bontis, "What's familiar is excellent: The impact of exposure effect on perceived journal quality," *Journal of Informetrics*, vol. 5, no. 1, pp. 219–223, Jan. 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1751157710000696 [Accessed: 2021-09-10]

[54] D. Sleeth-Keppler, "Familiarity as an Obstacle," May 2011. [Online]. Available: http://www.strategicbusinessinsights.com/about/featured/2011/2011-05-familiarityobstacle.shtml [Accessed: 2021-09-10]

[55] "Mere Exposure Effect - Biases & Heuristics," Jan. 2021. [Online]. Available: https://thedecisionlab.com/biases/mere-exposure-effect/ [Accessed: 2021-09-10]

[56] L. A. Zebrowitz, B. White, and K. Wieneke, "Mere Exposure and Racial Prejudice: Exposure to Other-Race Faces Increases Liking for Strangers of That Race," *Social cognition*, vol. 26, no. 3, pp. 259–275, 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2705986/ [Accessed: 2021-09-10]

[57] T. Cox Jr, "The multicultural organization," *Academy of Management Perspectives*, vol. 5, no. 2, pp. 34–47, 1991, publisher: Academy of Management Briarcliff Manor, NY 10510.

[58] K. A. Green, M. López, A. Wysocki, and K. Kepner, "Diversity in the workplace: Benefits, challenges, and the required managerial tools," *EDIS*, vol. 2002, no. 2, 2002.

[59] M. G. Fine, "Cultural Diversity in the Workplace: The State of the Field," *Journal of Business Communication*, vol. 33, no. 4, pp. 485–502, Oct. 1996, publisher: Association for Business Communication. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=bch&AN=9702175497&site=ehost-live [Accessed: 2021-09-13]

[60] D. Gaucher, J. Friesen, and A. C. Kay, "Evidence that gendered wording in job advertisements exists and sustains gender inequality." *Journal of Personality and Social Psychology*, vol. 101, no. 1, pp. 109–128, Jul. 2011. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0022530 [Accessed: 2022-01-13]

[61] E. Oldford and J. Fiset, "Decoding bias: Gendered language in finance internship job postings," *Journal of Behavioral and Experimental Finance*, vol. 31, p. 100544, Sep. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214635021000885 [Accessed: 2021-09-13]

[62] B. Gatzert, "A Toolkit for Recruiting and Hiring A More Diverse Workforce," 2013. [Online]. Available: https://www.unitedwaylima.org/sites/unitedwaylima.org/files/2022-03/A%20Toolkit%20for%20Recruiting%20and%20Hiring%2C%20Tang%20Cntr.pdf

[63] E. H. Chang, K. L. Milkman, L. J. Zarrow, K. Brabaw, D. M. Gromet, R. Rebele, C. Massey, A. L. Duckworth, and A. Grant, "Does Diversity Training Work the Way It's Supposed To?" *Harvard Business Review*, Jul. 2019, section: Diversity and inclusion. [Online]. Available: https://hbr.org/2019/07/does-diversity-training-work-the-way-its-supposed-to [Accessed: 2021-09-13]

[64] M. Noon, "Pointless Diversity Training: Unconscious Bias, New Racism and Agency," *Work, Employment and Society*, vol. 32, no. 1, pp. 198–209, Feb. 2018, publisher: SAGE Publications Ltd. [Online]. Available: https://doi.org/10.1177/0950017017719841 [Accessed: 2021-08-09]

[65] F. Dobbin and A. Kalev, "Why Diversity Programs Fail," *Harvard Business Review*, Jul. 2016, section: Human resource management. [Online]. Available: https://hbr.org/2016/07/why-diversity-programs-fail [Accessed: 2021-09-13]

[66] U. Rinne, "Anonymous job applications and hiring discrimination," *IZA World of Labor*, Oct. 2018. [Online]. Available: https://wol.iza.org/articles/anonymous-job-applications-and-hiring-discrimination/long [Accessed: 2021-01-10]

[67] I. Bohnet, *What Works: Gender Equality by Design*. Harvard University Press, Mar. 2016. [Online]. Available: https://hup.degruyter.com/view/title/521191 [Accessed: 2021-01-12]

[68] M. J. Hiscox, T. Oliver, M. Ridgway, L. Arcos-Holzinger, A. Warren, and A. Willis, "Going blind to see more clearly: Unconscious bias in Australian Public Service shortlisting processes," *Behavioural Economics Team of the Australian Government*, 2017.

[69] A. Krause, U. Rinne, and K. F. Zimmermann, "Anonymous job applications in Europe," *IZA Journal of European Labor Studies*, vol. 1, no. 1, p. 5, Dec. 2012. [Online]. Available: https://doi.org/10.1186/2193-9012-1-5 [Accessed: 2021-01-10]

[70] I. Kotlyar, "The effect of nonverbal cues on relationship formation," *Computers in Human Behavior*, p. 8, 2013.

[71] F. B. Mandal, "Nonverbal Communication in Humans," *Journal of Human Behavior in the Social Environment*, vol. 24, no. 4, pp. 417–421, May 2014. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10911359.2013.831288 [Accessed: 2022-07-14]

[72] N. S. Baron, *Always on: Language in an online and mobile world*. Oxford University Press, 2010.

[73] D. Derks, A. E. R. Bos, and J. von Grumbkow, "Emoticons and Online Message Interpretation," *Social Science Computer Review*, vol. 26, no. 3, pp. 379–388, Aug. 2008. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0894439307311611 [Accessed: 2022-07-19]

[74] A. H. Huang, D. C. Yen, and X. Zhang, "Exploring the potential effects of emoticons," *Information & Management*, vol. 45, no. 7, pp. 466–473, Nov. 2008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S037872060800089X [Accessed: 2022-07-19]

[75] J. B. Walther and K. P. D'Addario, "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication," *Social Science Computer Review*, vol. 19, no. 3, pp. 324–347, Aug. 2001. [Online]. Available: http://journals.sagepub.com/doi/10.1177/089443930101900307 [Accessed: 2024-04-04]

[76] J. B. Walther, "Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective," *Communication Research*, vol. 19, no. 1, pp. 52–90, Feb. 1992. [Online]. Available: http://journals.sagepub.com/doi/10.1177/009365092019001003 [Accessed: 2024-04-04]

[77] A. C. High and S. E. Caplan, "Social anxiety and computer-mediated communication during initial interactions: Implications for the hyperpersonal perspective," *Computers in Human Behavior*, vol. 25, no. 2, pp. 475–482, Mar. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563208002021 [Accessed: 2022-07-14]

[78] J. B. Walther, "Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction," *Communication Research*, vol. 23, no. 1, pp. 3–43, Feb. 1996. [Online]. Available: http://journals.sagepub.com/doi/10.1177/009365096023001001 [Accessed: 2022-08-09]

[79] D. L. Stone, D. L. Deadrick, K. M. Lukaszewski, and R. Johnson, "The influence of technology on the future of human resource management," *Human Resource Management Review*, vol. 25, no. 2, pp. 216–231, Jun. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053482215000030 [Accessed: 2021-06-21]

[80] J. S. Black and P. v. Esch, "AI-enabled recruiting: What is it and how should a manager use it?" *Business Horizons*, vol. 63, no. 2, pp. 215–226, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0007681319301612

[81] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–30, Nov. 2019. [Online]. Available: https://dl.acm.org/doi/10.1145/3359301 [Accessed: 2021-06-25]

[82] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters*, Oct. 2018. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G [Accessed: 2021-06-25]

[83] A. A. Kodiyan, "An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool," *Researchgate Preprint*, pp. 1–19, Dec. 2019.

[84] D. Ghosh, "AI is the future of hiring, but it's far from immune to bias," *Quartz*, p. 7, Oct. 2017. [Online]. Available: https://qz.com/work/1098954/ai-is-the-future-of-hiring-but-it-could-introduce-bias-if-were-not-careful/

[85] D. Allhutter, F. Cech, F. Fischer, G. Grill, and A. Mager, "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," *Frontiers in Big Data*, vol. 3, p. 5, Feb. 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fdata.2020.00005/full [Accessed: 2021-09-15]

[86] G. LLC., "Google Diversity Annual Report 2020," Google LLC., Mountain View, CA 94043, USA, Corporate Diversity Report, 2020. [Online]. Available: https://about.google/belonging/diversity-annual-report/2021/static/pdfs/google_2020_diversity_annual_report.pdf [Accessed: 2023-04-13]

[87] Textio, "Textio | Augmented writing for supercharged hiring," 2021. [Online]. Available: https://textio.com/products/ [Accessed: 2022-01-20]

[88] L. Gannes, "Textio Spell Checks for Gender Bias," Apr. 2015. [Online]. Available: https://www.vox.com/2015/4/20/11561646/textio-spell-checks-for-gender-bias [Accessed: 2021-09-16]

[89] L. Yarger, P. F. Cobb, and B. Neupane, "Algorithmic equity in the hiring of underrepresented IT job candidates," *Online Information Review*, vol. 44, no. 2, pp. 383–395, Jan. 2019, publisher: Emerald Publishing Limited. [Online]. Available: https://doi.org/10.1108/OIR-10-2018-0334 [Accessed: 2021-12-15]

[90] M. I. of Technology, "The MIT License | Open Source Initiative," 1987. [Online]. Available: https://opensource.org/licenses/MIT [Accessed: 2022-01-20]

[91] Datapeople, "Smart Job Post Editor for Inclusive Job Descriptions," Apr. 2024. [Online]. Available: https://datapeople.io/language-analytics-for-job-posts/ [Accessed: 2024-04-29]

[92] T. S. Mohr, "Why Women Don't Apply for Jobs Unless They're 100% Qualified," *Harvard Business Review*, Aug. 2014, section: Gender. [Online]. Available: https://hbr.org/2014/08/why-women-dont-apply-for-jobs-unless-theyre-100-qualified [Accessed: 2022-01-15]

[93] Ongig, "Text Analyzer for Job Descriptions," 2021. [Online]. Available: https://www.ongig.com/text-analyzer [Accessed: 2022-01-16]

[94] Pinpoint, "Improve Diversity & Inclusion in Recruitment," 2021. [Online]. Available: https://www.pinpointhq.com/diversity-recruitment-software/ [Accessed: 2022-01-16]

[95] Blendoor, "Blendoor," 2021. [Online]. Available: https://blendoor.com/ [Accessed: 2022-01-20]

[96] R. Siegel, "Blendoor App Breaks Down Computer Bias In Hiring," *NPR*, Mar. 2016. [Online]. Available: https://www.npr.org/2016/03/14/470427617/blendoor-app-breaks-down-computer-bias-in-hiring [Accessed: 2021-09-16]

[97] Entelo, "Diversity & Inclusion Recruiting Software | Entelo," 2021. [Online]. Available: https://www.entelo.com/products/platform/diversity [Accessed: 2022-01-16]

[98] Talvista, "Redacted Resume Screening for Objective Reviews," 2022. [Online]. Available: https://www.talvista.com/redacted-resume-reviews/ [Accessed: 2022-01-20]

[99] T. Hire, "Hiring Tests Boost Diversity and Inclusion," 2021. [Online]. Available: https://toggl.com/hire/pages/diversity-and-inclusion [Accessed: 2022-01-18]

[100] J. Dijkkamp, "The recruiter of the future, a qualitative study in AI supported recruitment process," Ph.D. dissertation, University of Twente, Enschede, Nov. 2019.

[101] M. Raub, "Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices," *ARKANSAS LAW REVIEW*, vol. 71, p. 43, Dec. 2018.

[102] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Barcelona Spain: ACM, Jan. 2020, pp. 469–481. [Online]. Available: https://dl.acm.org/doi/10.1145/3351095.3372828 [Accessed: 2021-12-15]

[103] S. Fazelpour and Z. C. Lipton, "Algorithmic Fairness from a Non-ideal Perspective," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. New York NY USA: ACM, Feb. 2020, pp. 57–63. [Online]. Available: https://dl.acm.org/doi/10.1145/3375627.3375828 [Accessed: 2021-12-15]

[104] N. Lewis, "Will AI remove hiring bias," *SHRM*, 2018.

[105] Pymetrics, "Case Studies," 2022. [Online]. Available: https://www.pymetrics.ai/case-studies [Accessed: 2022-01-18]

[106] N. Leong, "The Race-Neutral Workplace of the Future," *U.C. Davis Law Review*, vol. 51, no. 2, pp. 719–730, Dec. 2017.

[107] A. Lerner, "You can't fix diversity in tech without fixing the technical interview." Aug. 2018. [Online]. Available: https://medium.com/diversity-together/you-cant-fix-diversity-in-tech-without-fixing-the-technical-interview-597250e8564d [Accessed: 2022-01-18]

[108] M. Behroozi, C. Parnin, and T. Barik, "Hiring is Broken: What Do Developers Say About Technical Interviews?" in *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Oct. 2019, pp. 1–9, iSSN: 1943-6106.

[109] Applied, "Applied | Recruitment Platform | Recruitment Software," 2021. [Online]. Available: https://www.beapplied.com/ [Accessed: 2022-01-20]

[110] A. Blandford, D. Furniss, and S. Makri, "Qualitative HCI Research: Going Behind the Scenes," *Synthesis Lectures on Human-Centered Informatics*, vol. 9, no. 1, pp. 1–115, Apr. 2016. [Online]. Available: http://www.morganclaypool.com/doi/10.2200/S00706ED1V01Y201602HCI034 [Accessed: 2021-10-14]

[111] P. Mayring and T. Fenzl, "Qualitative Inhaltsanalyse," in *Handbuch Methoden der empirischen Sozialforschung*, N. Baur and J. Blasius, Eds. Wiesbaden: Springer Fachmedien, 2019, pp. 633–648. [Online]. Available: https://doi.org/10.1007/978-3-658-21308-4_42 [Accessed: 2021-08-20]

[112] P. Mayring, *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt: Beltz, 2014.

[113] K. Dali, ""Culture Fit" as "Anti-Diversity": Avoiding Human Resources Decisions that Disadvantage the Brightest," *The International Journal of Information, Diversity, & Inclusion (IJIDI)*, vol. 2, no. 4, Nov. 2018. [Online]. Available: https://jps.library.utoronto.ca/index.php/ijidi/article/view/32199 [Accessed: 2024-04-06]

[114] D. A. Norman and S. W. Draper, *User Centered System Design; New Perspectives on Human-Computer Interaction*. USA: L. Erlbaum Associates Inc., 1986.

[115] S. Highhouse, "Stubborn Reliance on Intuition and Subjectivity in Employee Selection," *Industrial and Organizational Psychology*, vol. 1, no. 3, pp. 333–342, Sep. 2008. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1754942600000584/type/journal_article [Accessed: 2021-09-24]

[116] A. Consulting, "Pyramid." [Online]. Available: https://trypyramid.com/ [Accessed: 2024-04-30]

[117] SQLAlchemy, "SQLAlchemy." [Online]. Available: https://www.sqlalchemy.org [Accessed: 2024-04-30]

[118] Pallets, "Jinja2." [Online]. Available: https://palletsprojects.com/p/jinja/ [Accessed: 2024-04-30]

[119] SQLite, "SQLite." [Online]. Available: https://www.sqlite.org/ [Accessed: 2024-04-30]

[120] P. S. Foundation, "Python," Apr. 2024. [Online]. Available: https://www.python.org/ [Accessed: 2024-04-30]

[121] M. Otto and J. Thornton, "Bootstrap." [Online]. Available: https://getbootstrap.com/ [Accessed: 2024-04-30]

[122] Converse, "Converse.js." [Online]. Available: https://conversejs.org/ [Accessed: 2024-04-30]

[123] X. S. Foundation, "XMPP." [Online]. Available: https://xmpp.org/ [Accessed: 2024-04-30]

[124] Strophe.js, "Strophe.js." [Online]. Available: https://strophe.im/strophejs/ [Accessed: 2024-04-30]

[125] D. Inc., "Docker," May 2022. [Online]. Available: https://www.docker.com/ [Accessed: 2024-04-30]

[126] K. Alvefur and M. Wild, "Prosody IM," last Modified: 2023-09-06. [Online]. Available: https://prosody.im/index [Accessed: 2024-04-30]

[127] F. Inc., "NGINX." [Online]. Available: https://www.nginx.com/ [Accessed: 2024-04-30]

[128] R. Garland, "The Mid-Point on a Rating Scale: Is it Desirable?" *Marketing Bulletin*, vol. 2, no. 1, pp. 66–70, 1991.

[129] C. Dancey and J. Reidy, *Statistics Without Maths for Psychology*, 8th ed. Pearson, 2020.

[130] L. A. Rivera, "Employer Decision Making," *Annual Review of Sociology*, vol. 46, no. 1, pp. 215–232, Jul. 2020. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-soc-121919-054633 [Accessed: 2024-03-21]

[131] M. Minsky, "A framework for representing knowledge," 1974.

[132] S. Cheryan and H. R. Markus, "Masculine defaults: Identifying and mitigating hidden cultural biases." *Psychological Review*, vol. 127, no. 6, pp. 1022–1052, Nov. 2020. [Online]. Available: https://doi.apa.org/doi/10.1037/rev0000209 [Accessed: 2024-03-28]

[133] E. Makarova, B. Aeschlimann, and W. Herzog, "The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students' Career Aspirations," *Frontiers in Education*, vol. 4, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2019.00060 [Accessed: 2022-06-21]

[134] S. Kruger and B. Hermann, "Can an Online Service Predict Gender? On the State-of-the-Art in Gender Identification from Texts," in *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)*. Montreal, QC, Canada: IEEE, May 2019, pp. 13–16. [Online]. Available: https://ieeexplore.ieee.org/document/8819508/ [Accessed: 2022-07-19]

[135] C. Fletcher, "Ethical Issues in the Selection Interview," *Journal of Business Ethics*, vol. 11, no. 5/6, pp. 361–367, 1992, publisher: Springer. [Online]. Available: https://www.jstor.org/stable/25072285 [Accessed: 2022-06-21]

[136] J. McCarthy and R. Goffin, "Measuring Job Interview Anxiety: Beyond Weak Knees and Sweaty Palms," *Personnel Psychology*, vol. 57, no. 3, pp. 607–637, Sep. 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1744-6570.2004.00002.x [Accessed: 2022-06-21]

[137] R. D. Goodwin, A. H. Weinberger, J. H. Kim, M. Wu, and S. Galea, "Trends in anxiety among adults in the United States, 2008–2018: Rapid increases among young adults," *Journal of Psychiatric Research*, vol. 130, pp. 441–446, Nov. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0022395620309250 [Accessed: 2024-03-22]

[138] R. Kindred and G. Bates, "The Influence of the COVID-19 Pandemic on Social Anxiety: A Systematic Review," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2362, Jan. 2023. [Online]. Available: https://www.mdpi.com/1660-4601/20/3/2362 [Accessed: 2024-03-22]

[139] J. G. Shalom, H. Israeli, O. Markovitzky, and J. D. Lipsitz, "Social anxiety and physiological arousal during computer mediated vs. face to face communication," *Computers in Human Behavior*, vol. 44, pp. 202–208, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563214006669 [Accessed: 2022-06-21]

[140] A. N. Finnerty, S. Muralidhar, L. S. Nguyen, F. Pianesi, and D. Gatica-Perez, "Stressful first impressions in job interviews," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. Tokyo Japan: ACM, Oct. 2016, pp. 325–332. [Online]. Available: https://dl.acm.org/doi/10.1145/2993148.2993198 [Accessed: 2022-06-21]

[141] K. G. Melchers, U.-C. Klehe, G. M. Richter, M. Kleinmann, C. J. König, and F. Lievens, ""I Know What You Want to Know": The Impact of Interviewees' Ability to Identify Criteria on Interview Performance and Construct-Related Validity," *Human Performance*, vol. 22, no. 4, pp. 355–374, Sep. 2009. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/08959280903120295 [Accessed: 2024-04-05]

[142] M. Kleinmann, P. V. Ingold, F. Lievens, A. Jansen, K. G. Melchers, and C. J. König, "A different look at why selection procedures work: The role of candidates' ability to identify criteria," *Organizational Psychology Review*, vol. 1, no. 2, pp. 128–146, May 2011. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2041386610387000 [Accessed: 2024-04-05]

[143] R. E. Rice and G. Love, "Electronic emotion: Socioemotional content in a computer-mediated communication network," *Communication research*, vol. 14, no. 1, pp. 85–108, 1987, publisher: Sage Newbury Park.

[144] P. Taesler and M. Janneck, "Emoticons und Personenwahrnehmung: Der Einfluss von Emoticons auf die Einschätzung unbekannter Kommunikationspartner in der Online-Kommunikation," *Gruppendynamik und Organisationsberatung*, vol. 41, no. 4, pp. 375–384, Dec. 2010. [Online]. Available: http://link.springer.com/10.1007/s11612-010-0124-5 [Accessed: 2022-07-19]