BIAS

Mitigating biases of AI in the labour market

Deliverable 2.1

The intermediate report of the mapping, survey and expert interviews

Contact us

info@biasproject.eu
www.biasproject.eu

Funded by the European Union

@BIASProjectEU

BIAS

Mitigating biases of AI in the labour market

Deliverable 2.1

The intermediate report of the mapping, survey and expert interviews

Deliverable type

Report

Month and date of delivery

28 July 2023

Work package

2

Leader

Leiden University (ULEID)

Dissemination level

Public

Authors

Carlotta Rigotti, Eduard Fosch-Villaronga

Programme	Contract Number	Duration	Start
Horizon Europe	101070468	48	November 2022

Funded by the European Union

email info@biasproject.eu

website www.biasproject.eu

Contributors

Name	Organisation
Carlotta Rigotti	ULEID
Eduard Fosch-Villaronga	ULEID
Guðbjörg Linda Rafnsdóttir	HI

Peer Reviews

Name	Organisation
Maria Sangiuliano	SVEN
Marzia Cescon	SVEN
Guðbjörg Linda Rafnsdóttir	HI

Revision History

Version	Date	Author(s), reviewer(s)	Modifications
0.1	14/06/2023	Carlotta Rigotti (ULEID)	First draft
0.2	30/06/2023	Guðbjörg Linda Rafnsdóttir (HI)	Internal review with minor revisions
0.3	05/07/2023	Maria Sangiuliano, Marzia Cescon (SVEN)	Internal review with minor revision
0.4	11/07/2023	Carlotta Rigotti (ULEID)	Second draft
0.5	18/07/2023	Eduard Fosch-Villaronga (ULEID)	Final draft for submission
1.0	28/07/2023	Roger Søraa, Mark Kharas (NTNU)	Final coordinator review with minor edits

Partner abbreviations	Full name
NTNU	Norwegian University of Science and Technology
HI	Háskóli Íslands (University of Iceland)
LOBA	Loba
CHX	Crowd Helix
SVEN	Smart Venice
DIGI	Digiotouch
ULEID	Leiden University
FARPL	Farplas Automotive
BFH	Bern University of Applied Sciences

Abbreviation	Meaning
AI	Artificial intelligence
The AI Act	The proposal for a European regulation laying down harmonized rules on artificial intelligence
AI HLEG	High-Level Expert Group on Artificial Intelligence
ALTAI	The Assessment List for Trustworthy Artificial Intelligence
DPAs	Data protection authorities
DPIAs	Data protection impact assessments
The FAccT community	The fairness, accountability, and transparency community
GDPR	General Data Protection Regulation
HR	Human resources
The Platform Work Directive	The proposal for a Directive on improving the working conditions in platform work
WP	Work package

Table of contents

Funded by the European Union

email
info@biasproject.eu
website
www.biasproject.eu

Funded by the European Union

email
info@biasproject.eu
website
www.biasproject.eu

Table of figures

Funded by the European Union

email
info@biasproject.eu

1 Executive Summary

This deliverable explains how the BIAS Consortium has engaged in extensive desk and empirical research in consultation with a diverse pool of stakeholders. At the heart of this process lies the aim to gain consolidated knowledge about AI applications and fairness in the labor market and lay down the basis for the design of the Debiaser and other project activities (e.g., the performance of the ALTAI (the Assessment List for Trustworthy Artificial Intelligence)).

After a short introduction, Section 3 begins with a scoping review that provides a state of evidence in the field of AI-driven recruitment and selection procedures and discusses competing formulations of fairness and bias from a philosophical, social science, and legal angle. The scope is then narrowed down to the recruitment and selection process and the ambitions and limitations of legal and, to a minor extent, extra-legal measures to ensure fairness and address diversity biases.

Section 4 explains the creation and purpose of the national labs, namely the creation of a pool of diverse stakeholders (e.g., employees, employers, HR practitioners, AI specialists, policymakers, trade union representatives, representatives of civil-based society organizations, and scholars) that could contribute or be interested in the implementation of the BIAS project, because of their professional expertise and experience. Section 5 covers the mapping exercise, which seeks to scope the current design and use of AI applications for HR management in Estonia, Iceland, Italy, The Netherlands, Norway, and Türkiye. A significant lack of transparency and compliance with data protection law is observed.

Section 6 summarizes the expert interviews, where the BIAS Consortium focused on the professional experiences and personal attitudes of HR executives and AI developers towards diversity bias of AI applications for hiring and management purposes. Major emphasis is put on creating, identifying, and mitigating diversity bias, with gender and race being considered the most common grounds for discrimination. Section 7 provides a preliminary analysis of the research outputs of the survey designed to reveal the fairness perceptions of job applicants and workers when interacting with AI applications in the labor market. In this context, fairness is mostly seen through procedural and non-discrimination lenses. A severe lack of transparency is likewise reported, while age and education seemingly are common grounds for discrimination.

The deliverable finishes with conclusions providing a bridge between the different project activities while creating new synergy with other WPs. The final version of this document is due in M15, namely January 2024.

Funded by the European Union

2 Introduction

The last decades have seen a growing trend toward the design and deployment of artificial intelligence (from now on AI) applications in the labor market, especially for recruitment and selection purposes. Given their ability to process information and make decisions at volumes and speeds that far exceed human capacity, AI applications could more effectively identify, attract, screen, assess, interview, and coordinate with job applicants. Nonetheless, the ambitions of AI applications for recruitment and selection purposes carry several limitations. Notably, there is increasing evidence that AI applications reproduce and perpetuate diversity bias and, as such, discriminate against job applicants due to their personal characteristics. Accordingly, if HR practitioners and employers plan to increasingly benefit from the advantages of technological innovation, they must use caution in identifying and mitigating the related risks.

The BIAS project aims to help HR practitioners and employers with this endeavor by developing a trustworthy and innovative technology based on natural language processing and cased based reasoning for use in a recruitment and selection process (from now on, the Debiaser). For this purpose, the BIAS Consortium is engaging in extensive consultation and co-creation with a diverse pool of stakeholders to gain comprehensive and nuanced knowledge about the state-of-the-art AI applications in the labor market and manifold conceptual understandings of fairness and diversity biases in this field. In doing so, the BIAS Consortium acknowledges that recruitment and selection decisions could have a significant and long-lasting impact on the job applicant's dignity, autonomy, and well-being since they are inextricably intertwined with their economic situation, housing possibilities, dignity, and well-being.

The overall structure of this Deliverable takes the form of 8 sections, the executive summary and this introduction being the first two. Each section reflects a different project activity, which varies according to methodology and target group but aims at gaining consolidated knowledge about fairness and diversity bias of AI applications in the labor market, especially regarding AI-driven recruitment and selection processes. On such premises, Section 3 begins with the cross-disciplinary literature review. ULEID focused on the reasons, ambitions, and limitations of AI-driven recruitment and selection processes and discussed some current ways to identify and mitigate the related diversity bias eventually. Section 4 explains the creation of the national labs, that is to say, a pool of stakeholders based in the country of origin of the Consortium partners and expected to engage and remain updated about most project activities. Section 5 covers the research findings of the mapping exercise, where most Consortium partners tried to examine some AI applications already in use in their own country. Section 6 reports the main research findings of the expert interviews targeting the professional experience and personal attitude of HR practitioners and AI developers towards diversity biases of AI applications in the labor market. Section 7 provides a brief overview of the ongoing survey. It seeks to map past or ongoing interactions between workers or job applications and AI applications, particularly regarding their perception of fairness and discrimination. In Section 8, the deliverable finishes with some preliminary discussion and conclusion that will be updated and expanded in month 15 (i.e., January 2024), with the submission of the final report (i.e., Deliverable 2.3).

3 The literature review

The literature review lays down the basis of the present deliverable in the sense that it examines the evolution and state-of-the-art research on AI applications in the labor market in order to provide further direction for future project activities, especially in the context of the other tasks of Work Package (from now on WP) 2, as well as the deployment of the De-biaser in WP3. Particular emphasis is put on the ambitions and limitations of AI applications in the recruitment and selection process, as well as divergent perceptions of their fairness.

3.1 Methodology

For the present deliverable, ULEID decided to opt for a scoping review, for providing an overview of the state of evidence in the field of AI applications for recruitment and selection purposes, particularly regarding fairness and diversity bias. The scoping review aims to 1) scope a body of literature, 2) clarify critical concepts, and 3) identify knowledge gaps.

This scoping review started in November 2022 and was divided into two iterations. In the first iteration (November 2022 - January 2023), ULEID conducted the scoping review with the support of all the Consortium partners but LOBA and Crowdhelix. For this purpose, we drafted guidelines to find helpful source materials for the research by including guiding research questions focusing, inter alia, on AI impact, AI limitations and ambitions, legal, ethical, and social issues, definition, identification, and mitigation of diversity biases (Annex I). Also, we created a shared folder on the BIAS Microsoft Teams platform, where partners could upload the documents and indicate the primary information about the article (i.e., complete reference, DOI, keywords, abstract, and additional remarks). A shared library on Zotero is also currently available to the Consortium partners. Thanks to the Consortium partners' different expertise, it was possible for ULEID to perform a cross-disciplinary literary review that currently ranges over the law, psychology, computer science, and labor studies, to name just a few examples. The second iteration (February 2023-April 2023) was performed by ULEID and sought to fill remaining knowledge gaps and examine in further detail issues that had arisen in the meanwhile during the performance of other WP2 project activities.

3.2 The research findings

The scope of the literature on fairness and diversity biases of AI applications in the labor market is potentially borderless, meaning that ULEID had to narrow it down based on the purpose of the present deliverable, namely, gain consolidated knowledge that could help the BIAS technical partners design the Debiaser in WP3 and inform other WP2 project activities, especially the expert interview and the survey. Accordingly, ULEID decided to focus on the ambitions and limitations of AI applications for recruitment and selection purposes, with particular regard to defining and addressing fairness and diversity bias.

3.2.1 The deployment of AI applications in the labor market: Its ambitions and limitations

While the deployment of AI applications in the labor market has made quite an impact on several HR practices (Ajunwa, 2020) (Tambe et al., 2019) (Raub McKenzie, 2018) (Black & van Esch, 2020) (Ebert et al., 2021), the scope of this literature review primarily lies in the recruitment and selection process since it is the main focus of this Work Package, as well as its links to WP3 and, more precisely, the design of the Debiaser.

Until the late 1990s, the HR practice of recruiting and selecting new personnel was done manually. HR practitioners were expected to attract job applicants, file and screen applications, and select who should move on to the next step of the hiring process or already be hired. The analog nature of the recruitment and selection process made it time and energy-consuming while also showing the conscious and unconscious biases of the HR practitioners.

With the launch and rapid spread of the Internet in the 1990s, the HR practice of recruitment and selection was brought to the next level. Digital job boards, for example, started to collect and show many job vacancies while targeting and reaching many potential job applicants at minimal cost and in a more appealing way. In this scenario, the network effect was exponential in that the more job vacancies websites could show, the more job seekers they could attract, and the more job seekers they could attract, the more job vacancies they could nudge employees to advertise. Furthermore, cyber recruitment took the form of professional network platforms, which allowed people to develop and grow as a community of work-related interest, exchange, information, and endorsement. This is the case, for instance, of LinkedIn (Black & van Esch, 2020)(Autor, 2001)(Peter Cappelli, 2001)(Sharone, 2017).

At present, HR practitioners are trying to keep pace with technological innovation and are increasingly resorting to AI-based recruitment and selection methods (Albert, 2019). More precisely, the design and deployment of AI applications in recruitment and selection processes attempt to facilitate or replace HR practitioners in performing four tasks: outreach, screening, assessment, and coordination (Black & van Esch, 2020). This means that AI applications could assist HR practitioners with identifying job applicants and pooling the best possible ones by learning and placing job vacancies via banners, popups, emails, and texts for optimal uptake and response. Besides, AI applications could be better than people when screening job applications because they not only take less time but can also deduce specific skills and personality traits from the digital records of the job applicant, such as their social media footprint. Lastly, AI applications could improve the possibilities for assessment, with gamification being a classic example, as well as take the coordination amongst the various stages of the recruitment and selection process over (Black & van Esch, 2020) (Naim et al., 2018) (Gonzalez et al., 2022).

When saying that AI applications are designed and deployed to facilitate or replace HR practitioners, the latter either continues to lead and manage the recruitment and selection process that the technology facilitates rather than leaving all the tasks to the technology. Even the most developed AI-driven recruitment and selection process has some degree of human involvement like in developing the technology, deciding which datasets to feed into it, or supervising their training and testing (Gonzalez et al., 2022).

Overall, the recruitment and selection of the right person are crucial to the survival and success of an organization because it contributes to pooling the best knowledge, skills, abilities, and other characteristics needed for competitive advantage (Acikgoz et al., 2020, p. 399) (Black & van Esch, 2020) (Koivunen et al., 2019). For Eva Derous and Filip De Fruyt, the growing deployment of AI applications for recruitment and selection purposes arises from a strategic impetus following the so-called 'war for talent' that organizations have recently engaged in (Derous & De Fruyt, 2016). It has been commonly assumed that AI applications can improve the efficiency of every recruitment and selection process, especially in terms of necessary time, money, and effort (Langer et al., 2021; M. et al., 2018)(Gonzalez et al., 2022)(Sánchez-Monedero et al., 2020) (Sousa & Wilks, 2018). Additionally, using AI applications can make the recruitment and selection process more flexible because it can occur without the time and place constraints (Langer et al., 2021; M. et al., 2018)(Gonzalez et al., 2022). Interestingly, Mohammad H. Jarrahi argues that technology-driven recruitment and selection of new personnel could help HR practitioners find a balance between the intuition and creativity of their human decision-making process and the analytical capacity to handle complex data offered by AI (Jarrahi, 2018). Furthermore, AI applications are seemingly better placed to identify the most critical criteria for matching a job applicant to a job vacancy (Koivunen et al., 2019). Lastly, since AI applications seem to perform recruitment and selection-related tasks without human involvement, it is sometimes believed that their decisions are less biased and should therefore be favored (van Esch et al., 2019) (Cowgill, 2019).

Against this backdrop, there is still a body of literature voicing some concerns and stressing the limitations of AI applications in recruitment and selection processes. Notably, the design and deployment of this technology are costly, meaning that big companies will always stay ahead of the competition, and small- and medium-sized enterprises are not likely to benefit from its advantages (Black & van Esch, 2020). Besides, for AI-driven recruitment and selection processes to work properly, they need to process personal and special categories of data, thereby giving rise to potential risks of privacy and data protection violations (Ajunwa, 2020), as well as personal experiences of social marginalization, stigmatization, and discrimination (Black & van Esch, 2020) (Chamorro-Premuzic et al., 2017) (Lee & Shin, 2020). This is extremely important, considering that AI applications might work and lead to recruitment and selection decisions that are complicated to understand fully (Gonzalez et al., 2022). Otherwise, technology might lack explicability and transparency, to the detriment of accountability (Sánchez-Monedero et al., 2020). Ultimately, HR practitioners sometimes oppose the deployment of AI applications for recruitment and selection purposes because they consider them a threat to their job rather than an aiding tool (Black & van Esch, 2020).

In the attempt to examine AI applications in recruitment and selection processes from every angle, a growing body of research looks at the reactions of job applicants, that is to say, all those feelings, attitudes, and behaviors arising from their personal experience with an AI-driven process of recruitment and selection. In this regard, empirical research offers contradictory findings about the perception of fairness and bias job applicants have when interacting with AI applications. On the one hand, Min Nicholas Folger *et al.* positively considers the potential of job applicants towards AI application. Job applicants, in fact, seemingly connect the use of AI applications with innovation and are, in turn, more attracted to the organization (Folger et al., 2022). On the other hand, Min K. Lee argues that job applicants perceive AI-driven recruitment and selection processes as less fair (Lee, 2018). Also, other studies report that job applicants complain about the impersonal treatment embedded in AI-driven recruitment and selection process and therefore feel demotivated to apply (Acikgoz et al., 2020) (Gonzalez et al., 2022). Another negative perception of job applications arising from deploying AI applications in recruitment and selection processes relates to a sense of unfamiliarity with the technology. This feeling could be inextricably intertwined with a lack of transparency in the way the technology is designed and deployed, where transparency is understood to cover the provision of information about the recruitment and selection process that makes it more predictable and justifiable (Gonzalez et al., 2022) (Langer et al., 2021) (Mirowska & Mesnet, 2022). Job applicants also raise privacy and data protection concerns, as well as fear of discrimination (Langer et al., 2021) (Acikgoz et al., 2020) (Langer et al., 2017). With special regard to professional network platforms - which are often AI-driven, such as LinkedIn- Ofer Sharone reports that this technology is considered double-edged, meaning that it could increase the job seeker's visibility and opportunities while exposing them to new vulnerabilities. It does not allow the individual to "customize the extent or nature of the disclosed information depending on the audience, the context, or the level of established trust" (Sharone, 2017).

In conclusion, it appears that AI applications are increasingly designed and used throughout the entire funnel of the recruitment and selection process. Nonetheless, this design and deployment carry with it several ambitions and limitations. Besides a claim for innovation and efficiency, it is commonly assumed that AI applications are likely to be biased. With this in mind, this literature review explores what counts as fairness and diversity bias of AI applications outside and within the labor market. The focus then shifts towards legal and, to some extent, non-legal measures that could ensure fairness and address diversity bias.

3.2.2 Generally defining fairness of AI applications: A first attempt

One major issue that has indisputably dominated any literature on the design and deployment of AI applications concerns the definition of fairness. In broad terms, Ninareh Mehrabi and colleagues (2019) argue that fairness is absent prejudice or favoritism toward an individual or a social group based on their inherent or acquired personal characteristics. As such, “technology is not merely a tool of implementation but constitutive of a [society]’s values commitments in practice” (Mulligan et al., 2019, p. 119), in the sense that every AI application is expected to be fair, by yielding its potential bias to fundamental values, principles, interests, and rights. This is also the case with the Debiaser that, in identifying and mitigating diversity bias of AI applications in the recruitment and selection process, can be considered fairness sensitive. However, because the BIAS Consortium does not assume that ‘fairness’ is a universal term, searching for a common definition is of the utmost importance given the design process. This subsection, therefore, tries to lay down the basis for this discussion and touches upon the main formulations of these concepts through cross-disciplinary lenses, along with some simple examples that make all these different formulations more understandable.[1]

From a philosophical angle, it is possible, to begin with the formal fairness, where the same rule is applied to everyone. For instance, the core of the ULEID team - part of the BIAS Consortium - comprises two people who have gained a Ph.D. in law. At first, this means that both people should be paid equally. However, one could argue that the two scholars should be paid

[1] The BIAS partners discussed the definition of fairness and diversity bias and their formulations during the Consortium meeting in Bern on 13 and 14 June 2023. Some of the joint conclusions are reported in the last section of this document. Additionally, NTNU and BFH are working on a separate document covering a more technical literature review and discussion on fairness. This draft is available on the MicrosoftTeams shared folder of the BIAS Consortium and is likely to turn into a peer-review article next year.

differently because of their professional experience or contribution to the research project. Also known as substantive fairness, this example proves that every formulation of fairness should vary and rely on the context or the details in a specific case, meaning that there is no free-standing conception of fairness that could serve as a common ground in this definitory search. In any case, this general distinction between formal and substantive fairness already traces back to Ancient Greek philosophy and can be found in the works of Plato and Aristotle (Aristotle et al., 2004). Since the IV century BC, a growing body of philosophical literature has emerged and has offered various definitions of fairness. For example, a name frequently associated with contemporary formulations of fairness is that of the philosopher John Rawls who describes a society of free citizens equally holding fundamental rights and cooperating within an egalitarian economic system, where they are aware that society works for everyone's benefit. Those lucky enough to be born with more significant natural potential are not getting better at the expense of those less lucky (Rawls & Kelly, 2001). In this context, fairness has a distributive flavor.

Within social science disciplines, every formulation of fairness is construed concerning a social asymmetry of power (Mulligan et al., 2019). While focusing on international trade, for instance, Steven Suranovic explains that "[t]o many people, the unequal distribution of income, wealth, and economic well-being is unfair […]. Consequently, policies seen as increasing the disparities […] are often judged to be unfair policies, while policies that reduce these inequalities are seen as fair" (Suranovic, 2010, p. 68). Instead, Nancy Fraser and Axel Honneth argue that, for society to be fair, it is necessary to be free from social prejudice and discrimination while drawing on the value of group differences and aiming at the redistribution of sources (Fraser & Honneth, 2003). Some authors even use the label 'fairness feminism' to indicate that part of scholarship and activism claims for women and men to be treated alike (Graham, 1988). On a different note, the concept of fairness could also relate to the environment, thereby covering respect for non-human creatures and future generations, the duty not to pollute, and social redistribution in economic growth (Dator, 2017).

This close link between the formulation of fairness and a social asymmetry of power is somewhat specific to the law, with fairness in international law relying on distributive justice and the right to process (Franck, 1998) to give an example. For this deliverable, however, it is worth anticipating two arguments: the definition of fairness in data protection and anti-discrimination laws. On the one hand, Article 5(1)(a) GDPR introduces the fundamental principle of fairness, lawfulness, and transparency. According to Gianclaudio Malgieri, the GDPR positions fairness within the substantial circumstances at stake to prevent unbalanced relationships between the data subject and other stakeholders (Malgieri, 2020). Within anti-discrimination law, on the other hand, it is possible to construe fairness in opposition to all those situations where a person is treated less favorably than another is, has been, or would be treated in a comparable situation due to their personal characteristics (*i.e.*, direct discrimination), or an apparently neutral provision, criterion or practice unjustifiably puts a person or some people at a particular disadvantage, in comparison with other people (*i.e.*, indirect discrimination) (FRA, 2018).

Finally, it is more likely for computer science and other scientific research to create their conceptual meanings of fairness by writing a line of code and an equation. In this regard, Mulligan et al. (2019) argue that fairness criteria are mostly developed according to a specific task or problem, such as fair division (also known as 'cake cutting') and voting in game theory. Another common distinction of fairness in computer science lies in individual fairness (*i.e.*, treating similar cases alike) and group fairness (*i.e.*, treating groups in a way that is somehow equal) (Bringas Colmenarejo et al., 2022)(Mulligan et al., 2019) (Hacker, 2018). With BFH and NTNU separately focusing on the technical formulations of fairness,[2] it is worth stressing that the computer science community increasingly uses fairness criteria to address structural subordination within society, in line with the said objectives specific to social sciences and law (Kasirzadeh, 2022)(Hanna et al., 2020). The design of the Debiaser illustrates this point clearly, by relying on an extensive consultation and co-creation process enshrined in WP2.

In the attempt to provide a bridge between the different research communities, the High-Level Expert Group on Artificial Intelligence (from now on AI HLEG) regards fairness as an ethical principle that is later transposed into a key requirement, be it a technical or non-technical method.[3] As an ethical principle, fairness is seen through substantial and procedural lenses. This means, on the one hand, that AI applications should ensure equal and just distribution of both benefits and costs while guaranteeing that each individual and group are free from unfair bias, discrimination, and stigmatization and could benefit from equal opportunities. Besides, using AI applications should never lead to people being deceived or unjustifiably impaired in their freedom of choice. On the other hand, procedural fairness implies the ability to contest and seek effective redress against decisions made by AI applications and the people operating them. For this purpose, whoever is accountable for the decision must be identifiable, and the decision-making processes should be explicable. In terms of critical requirements and its defining methods, the AI HLEG inextricably intertwines diversity and non-discrimination with fairness, in the sense that a fair AI application is expected to ensure equal access and treatment through an inclusive design process engaging with all those stakeholders who might directly or indirectly be affected through its life cycle (AI HLEG, 2019).

3.2.3 Defining diversity biases of AI applications: A first attempt

As with fairness, the definition of AI-driven bias has been a much-debated question in the last decades. David Danks and Alex J. London argue that the word 'bias' often has a negative connotation in the English language, with bias being something to avoid or necessarily problematic. However, the authors clarify that the term 'bias' should be better understood in an older, more neutral way, referring to deviation from a standard. This means that it is possible to have, for example, a statistical bias, where an estimate deviates from a statistical standard; a moral bias, where a judgment deviates from a moral norm, a legal bias, where conduct deviates from a legal provision; and so forth. Nonetheless, because there are as many types of bias as many types of standards being used, David Danks and Alex J. London highlight that the same AI application can be biased according to one standard but not the other (Danks & London, 2017). For example, an AI application favoring the selection and recruitment of women over men could abide by legal, ethical, and social norms, considering that it is well-established that women do not often participate in the workforce on equal footing. However, this preferential treatment

[2] See footnote no. 1.
[3] The European Commission established the AI HLEG in June 2018 to support its roadmap for regulating AI. The AI HLEG first launched a consultation process and published the 'Ethics Guidelines for Trustworthy Artificial Intelligence'. For more information, *see* https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

Funded by the European Union
email info@biasproject.eu
website www.biasproject.eu

could be considered an unfair statistical deviation from the overall population. Also, it involves making value-laden and normative decisions covering the distribution of employment opportunities and the factors affecting an individual's employment prospects. In any case, this deliverable draws on the understanding of bias suggested by Batya Friedman and Helen Nissenbaum. It refers to computer systems systematically and unfairly discriminating against a person or a group of people in favor of others (Friedman & Nissenbaum, 1996).

Batya Friedman and Helen Nissenbaum also divide bias into three groups. First, pre-existing bias arises from society and its constitutive elements, such as public institutions and private organizations. Also, they can reflect the personal bias of people designing, developing, or using the technology. This bias can be transposed through the explicit and conscious efforts of the developer(s) or the user(s) or implicitly and unconsciously. Second, technical bias comes to be understood as technical constraints or considerations deriving from the design process, such as ascribing social meaning to algorithms developed out of context and attempting to make social constructs amenable to computers. Third, emergent bias is specific to the context of the use of the technology, thereby arising after its design and from potential change in society in terms of values, principles, and interests (Friedman & Nissenbaum, 1996). Besides this three-fold classification, a growing body of literature has examined and reported different shapes and sources of bias in AI applications (Mehrabi et al., 2019) (Danks & London, 2017), including:

- **Historical bias** corresponds to human bias that exists worldwide and is transposed in the dataset, even given a perfect sampling and feature selection. In drafting the job vacancy for a CEO position, for example, an AI application could rely on the professional profile of previous people holding it, thereby discriminating against any job applicant with different age, gender, education, and other personal characteristics.
- **Representation bias** arises from the definition and sampling of the population, meaning that the absence of geographical diversity in datasets could demonstrate a bias toward third-country nationals, to give an example.
- **Measurement bias** is based on how the AI developer chooses, uses, and measures a particular feature. For instance, an AI application could use prior hospitalization of family members as a proxy variable to assess the employability of a job applicant.
- **Evaluation bias** occurs during model evaluation. In AI-driven job interviews, for example, the technology could use facial recognition systems biased toward skin pigmentation and eye shapes.
- **Aggregation bias** happens when false conclusions and assumptions are inferred for a social subgroup based on observing others. This is evident in the case of an AI-driven tool that monitors employees working remotely and disregards different care duties and other personal specificities in collecting biometric data (*e.g.*, eye shapes).
- **Sampling bias** arises from non-random sampling of subgroups, meaning that the trends estimated for the workplace of a given company – at a given time and in a given place - cannot generalize from data collected from another one.

No matter the shape and source of bias, the BIAS Consortium agrees with a growing body of literature that AI applications are designed and used by human beings (FRA, 2022; Harned & Wallach, 2020; Bendick & Nunes, 2012). This implies that those human beings make value judgments and assumptions about the design of AI applications and their functioning in the world, all of which are likely to be affected by their implicit or explicit bias. Also, it is worth stressing that the BIAS project only focuses on diversity bias, particularly gender- and race-related ones. Indeed, diversity in the workforce does not only abide by non-discrimination law at the EU level (FRA, 2018), but is also desirable because it can lead to social mobility for people with varying demographic characteristics and cultural and educational backgrounds, besides higher productivity, creativity, and critical skills in the workforce (Drosou et al., 2017).

3.2.4 Fairness of AI applications in the labor market

The last two subsections have demonstrated that a growing body of cross-disciplinary literature has proposed several formulations of fairness and bias. However, AI developers and users - who have to transpose all these definitions into practice - often warn against their unsuitability in real-world applications due to the possible misinterpretation, the increase in ethics washing, and the obfuscation of human accountability (Bringas Colmenarejo et al., 2022). To avoid these and other shortcomings, the BIAS Consortium, therefore, assumes that fairness and bias are situated knowledge and practices (Mulligan et al., 2019), with the following subsections focusing on how to operationalize them in the labor market and, more precisely, in the selection and recruitment process.

The International Labor Organization is the only international body promoting 'fair recruitment' as a new norm, which comes to be understood as any hiring taking place "in a way that respects, protects, and fulfills internationally recognized human rights, including those expressed in international labor standards, and in particular the right to freedom of association and collective bargaining, and prevention and elimination of forced labor, child labor, and discrimination in respect of employment and occupation" (Jones, 2022, pp. 312–313). At the EU level, fair recruitment relies on the actual decision-making process but not its result (Hauer et al., 2021). Said otherwise, the focus of anti-discrimination law does not lie on the hiring decision but on the treatment of the job applicant in the recruitment and selection process. Instead, in the literature, it is commonly held that, for a selection and recruitment procedure to be fair, it is necessary to match the job vacancy with the best job applicant, thereby answering the question of what counts as a good employee. However, the answer to this question is complicated and varies according to the stakeholder's viewpoint (Tambe et al., 2019), as well as the specific stage of the selection and recruitment process. Accordingly, what follows is a brief overview of these nuances in the literature covering both analog and AI-driven recruitment.

It appears that job applicants often perceive substantive fairness to derive from the desire for equitable treatment and outcome by comparing their knowledge, skills, and efforts with the hiring decision of the HR practitioner(s) (Zibarras & Patterson, 2015)(Schinkel et al., 2013) (Thorsteinson & Ryan, 1997) (Gilliland, 1993). Also, they connect substantive fairness with their personal experience of respect, dignity, and honesty during the selection and recruitment process, especially regarding human empathy and communication (Köchling & Wehner, 2023) (Alder & Gilbert, 2006). However, because most selection and recruitment processes are

Funded by the European Union

characterized by information asymmetry, job applicants are more likely to favor procedural fairness, where selection and recruitment procedures are considered a means to achieve fair hiring outcomes. In the literature (Mirowska & Mesnet, 2022) (Konradt et al., 2013) (Furnham & Chamorro-Premuzic, 2010) (Truxillo et al., 2004) (Van Vianen et al., 2004) (Gilliland et al., 2001) (Steiner & Gilliland, 2001)(Truxillo et al., 2001)(Van Den Bos et al., 1997) (Gilliland, 1993), the more common principles of procedural fairness are the following ones:

- **Job relatedness**: The recruitment and selection procedures should only assess the personal characteristics necessary for the job and can predict the skills and capabilities of the job applicant.
- **Consistency:** Each job applicant should undergo the same recruitment and selection process.
- **Opportunity to perform**: The job applicant can demonstrate their knowledge and skills during the recruitment and selection.
- **Objectivity**: Especially regarding collected information that should not, e.g., invade the job applicant's privacy.

Sometimes, the fairness perception of the job applicant depends on whether the recruitment and selection process is analog or AI-driven, with the latter being generally characterized by negative attitudes (Koch-Bayram et al., 2023) (Hilliard et al., 2022) (Wesche & Sonderegger, 2021). For Airlie Hilliard et al., this negative attitude derives from the less degree of control and influence that the job applicant believes to have (Hilliard et al., 2022). Irmela F. Koch-Bayram et al., instead, prove that people who more frequently experience social discrimination at work consider the deployment of AI applications to be able to increase fairness and organizational attractiveness (Koch-Bayram et al., 2023).

A search of the literature reveals few studies focusing on the fairness perception of the HR practitioner(s) in the selection and recruitment process, with G. Stoney Alder and Joseph Gilder, stressing the importance of the person-job fit (Alder & Gilbert, 2006). In other words, it is likely for the HR community to construe fairness in the selection and recruitment process concerning matching the knowledge, skills, abilities, and other characteristics of the job applicant with the core tasks, duties, and responsibilities of the job vacancy. This also implies that fairness arises from a selection and recruitment process that is the most significant benefit to the employer and the immediate stakeholders of the organization, irrespective of procedural fairness or any positive consequence on society (Alder & Gilbert, 2006). In this scenario, it is possible to imagine that HR practitioners do not only seek to identify the best job applicant in terms of knowledge, skills, and abilities but also consider whether their hiring decision fits within the overall organization (Roberson, 2013). Interestingly, Sami Koivunen *et al.* report that HR practitioners share the need for diversity in the workforce but agree that it is often practically challenging in the sense that it might collide with the practices or culture of a specific team or the overall organization (Koivunen et al., 2019). In any case, as differences in values may lead to differences in what they consider fair or not, Timothy E. Landon and Richard D. Arvey show that judgments about fairness likewise vary amongst HR practitioners (Landon & Arvey, 2007).

A growing body of literature argues that fairness should be considered a dynamic concept, with its perception evolving during the selection and recruitment process (Konradt et al., 2020)

(Butucescu & Iliescu, 2018) (Konradt et al., 2016). This means that a precise definition of fairness should draw on the specific stage of the selection and recruitment process as follows. Each recruitment and selection process typically begins with setting the recruitment objectives, in the sense that the HR practitioners describe the job vacancy requirements and often refer to previous employees. In this way, HR practitioners already identify the potential pool of job applicants and should avoid previous bias (*e.g.*, specific gender, education). In developing selection and recruitment strategies, HR practitioners identify the best communication and dissemination channels and draft job vacancies that draw the attention of desirable but diverse candidates. This involves, *e.g.*, employer branding, various recruitment platforms, and inclusive language. Particularly, AI applications are increasingly used to provide insight into the appeal of a job vacancy. Once the job vacancy is public, HR practitioners perform manifold selection and recruitment tasks, including storing and sorting job applications and pre-screening them.

Regarding pre-screening, AI applications could help HR practitioners eliminate unqualified applicants by evaluating the job applicants' knowledge, skills, and abilities. Later, they can facilitate the communication of the pre-screening results and the scheduling of further selection steps with the shortlisted candidates. In making recruitment decisions, the main aim is to identify the candidate who could be successful in a specific job position. Accordingly, the evaluation of each job applicant relies on parameters of 'person-job' and 'person-organization' fit. Based on the skills and abilities required to thrive in the job position, HR practitioners could use different assessment techniques, like interviews, personality tests, and work samples, that are increasingly AI-driven, as previously shown in Section 3.2.1.

Accordingly, the transposition of all these different fairness perceptions into practice is complex because it requires prioritization of numerous observable characteristics that make a job applicant a “good” one (Tambe et al., 2019) (McKenzie, 2018). This means that, to some extent, any hiring decision is primarily normative. However, considering also what has been previously said, it is possible to identify two main requirements that seemingly ensure a basic level of AI-driven fairness in the literature. On the one hand, the dataset should be as diverse, representative, and updated as possible (Gonzalez et al., 2022) (Kleinberg et al., 2018). On the other hand, the AI application should work in a transparent and explicable way, thereby allowing for collective scrutiny and verifiability (Kim, 2019) (Ananny & Crawford, 2018). Additionally, the research tries to satisfy these and other criteria and, as such, ensure AI-driven fairness by adopting legal and non-legal measures. Section 3.2.6 and ff. present some of them to position the design of the Debiaser within the contemporary state-of-the-art.

3.2.5 Diversity bias in the labor market

The notion of bias is far from being a novelty within the HR community, with Steward J. Black and Patrick van Esch reporting that, when the recruitment and selection process was still fully analog, HR practitioners used to be beset with cognitive biases that limited the reliability and validity of their judgments (Black & van Esch, 2020). Because, as previously said, the BIAS Consortium expects AI-driven bias to reflect the human ones, it is necessary to briefly summarize the most common bias in the selection and recruitment process (Harver, 2022) (Black & van Esch, 2020) (Barragán Díaz et al., 2019)(Equalture, 2023):

Funded by the European Union

- **Conscious versus unconscious bias**, where the latter means that most HR practitioners are likely not to be aware of their misleading assumptions, beliefs, or attitudes occur.
- **Anchoring bias**, where the first information the HR practitioner collects unduly affects how they interpret the following ones.
- **Confirmatory bias**, where the HR practitioner seeks out and observes information that confirms initial judgments – no matter whether they are positive or negative – about a job applicant.
- **Similarity bias**, where an HR practitioner unconsciously favors job applicants who are familiar to them, no matter whether those similarities are good predictors of their following job performance.
- **Stereotyping bias** is where the HR practitioner assumes a social group based on what they have learned about this social category.
- **Projection bias** is where the HR practitioner assumes that the job applicant thinks, behaves, and feels similarly to them, thereby wrongly anticipating their opinions, conduct, and feelings.
- **Attribution bias**, where the HR practitioner explains the job applicant's behavior by referring to their personal character rather than any situational factor. In doing so, they risk overestimating the weight of personality traits and underestimating the influence of their individual circumstances.
- Based on the **Halo effect**, the HR practitioner skips proper investigation of a job applicant's background and relies on one of their positive aspects, such as graduating from a prestigious university.
- Based on the **Horn effect**, the HR practitioner skips proper investigation of a job applicant's background and relies on one of their negative aspects, such as having some gap years on the CV.
- **Illusory correlation** occurs when the HR recruiter believes in a relationship between two variables that do not exist, such as asking the favorite animal to determine some soft skills.

3.2.6 Addressing diversity biases in AI applications in the labor market: Data protection law

Generally, AI applications process personal data. On the one hand, personal data might contribute to creating the dataset used to train the AI application. On the other hand, the AI application could be designed to process personal data, perform a specific task, and reach a specific goal. In the instant case, it is likely for each AI application for recruitment and selection purposes to be trained on personal data that had been collected from previous job vacancies, applications, and matchings and to examine then, categorize, score, or make decisions about the personal data of present job applicants. Accordingly, the design and deployment of these AI applications give rise to legal, ethical, and social concerns about the data protection and privacy of the individual, namely the legal, ethical, and social interest in lawful and proportionate processing of personal data subject to human oversight (Lagioia & Sartor, 2020). As already mentioned in Section 3.2.1, a growing body of literature voices concerns about the possible violation of the fundamental rights to privacy and data protection arising from the deployment of AI applications in the selection and recruitment process and leading to non-transparency, knowledge asymmetry, surveillance, and social subordination, to name just a few examples

(Aloisi, 2024) (Ebert et al., 2021) (Ebert et al., 2021) (Collins & Marassi, 2021) (Ajunwa, 2020). At the same time, although the General Data Protection Regulation (EU) 2016/679 (from now on, the GDPR) does not explicitly refer to AI applications, most of its provisions could be considered relevant to AI applications. They could address some diversity biases of AI applications in the labor market.

In brief, the GDPR aims to ensure the fundamental rights of the individual and facilitate the free movement of personal data within the European Union. For this purpose, the GDPR attributes several obligations to the data controller (*i.e.*, the natural or legal person determining the purposes and means of the processing of personal data) and several rights to the data subject (*i.e.*, the natural person who is identifiable via their personal data), while laying down some fundamental principles relating the data processing. With this in mind, some GDPR provisions could influence the design and development of AI applications in the labor market, especially regarding selection and recruitment purposes.

While Section 3.2.2 already discussed the formulation of fairness in data protection, Article 5 GDPR includes other fundamental principles whose compliance could help address diversity biases of AI applications at the selection and recruitment stage, namely lawfulness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality (Aloisi, 2024) (Hacker, 2018). This means, for example, that any output arising from a biased dataset is not accurate and violates Article 5(1)(d) GDPR or that each job applicant should be informed about the data processing done by an AI application in a concise, accessible, and understandable way based on Article 5(1)(a). In order to evaluate compliance with the principles enshrined in Article 5 GDPR, Antonio Aloisi refers to the performance of a data protection impact assessment (Aloisi, 2024).

According to Article 35(1) GDPR, it is necessary to perform a data protection impact assessment (from now on DPIA) when the processing of personal data could lead to a high risk to the rights and freedoms of natural persons, a high risk that is likely to stem from the design and deployment of AI applications for recruitment and selection purposes, due to their significant consequences on the dignity, autonomy, and well-being of the job applicant. In a nutshell, the performance of a DPIA involves identifying and evaluating the risks to the fundamental rights and freedoms of the individual to adopt a mitigation plan later. Accordingly, recent research has suggested that the performance of a DPIA could not only help the employee comply with the GDPR and prove the fairness of its AI application for recruitment and selection purposes but also demonstrate that *ex-ante* measures have been taken to identify and mitigate its possible diversity biases (Aloisi, 2024; Hacker, 2018).

Because it is likely for AI applications for recruitment and selection purposes to rate job applications and exclude some of them without human involvement, it has been commonly assumed that Articles 13-15 GDPR, as well as Article 22 GDPR, could likewise come to the rescue (Aloisi, 2024) (Parviainen, 2022) (Ebert et al., 2021) (Zuiderveen Borgesius, 2020) (Sánchez-Monedero et al., 2020). On the one hand, under Articles 13 and 14 of the GDPR, the right to information imposes an obligation to provide data subjects with "meaningful information about the logic involved, as well as the significance and envisaged consequences of such processing for [them]" (Aloisi, 2024)(Sánchez-Monedero et al., 2020). On the other hand, Article 22 of GDPR attributes to the data subject the right not to be subject to a decision "based solely on the automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Based on this interpretation, AI-driven hiring decisions are generally allowed. Only after a job applicant invokes their right can the related automated recruitment decision be restricted (Parviainen, 2022). Alternatively, the same provision is considered a general prohibition preventing the data controller from making automated decisions except for limited exceptions (Tosoni, 2021) (Bygrave, 2020). Indeed, Article 22(2) GDPR provides some exceptions that could legitimize automated decision-making in the recruitment and selection process, namely the necessity for entering or performing a contract and the individual's explicit consent. In this scenario, it is likely for automated decision-making to be considered necessary for entering an employment contract if it would be practically impossible or unreasonable to handle certain parts of the recruitment process without it by other less privacy- and rights-intrusive means. At the same time, for the explicit consent of the job applicant to be legitimate, the employer must make sure that it is freely given, informed, specific, and unambiguous and sees it in the context of the power asymmetry that typically characterizes its relationship with the job applicant. In any case, if the employer decides to deploy an automated decision-making procedure in compliance with the said exceptions, Article 22(3) GDPR still requires them to take "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision." This, according to Philipp Hacker, means that the absence of strategies to identify and mitigate diversity biases should amount to a violation of this provision (Hacker, 2018).

If there is a violation of all the GDPR provisions that have been so far mentioned, national data protection authorities (from now on DPAs) - who monitor and enforce the regulation according to Article 57(1)(a)GDPR -have a wide range of instruments at their disposal to enforce fairness (Hacker, 2018). For example, national DPAs could check the existence of bias identification and mitigation strategies by requesting all relevant information, demanding access to personal data, and conducting audits, based on Article 58(1) GDPR. Also, they could enforce bias identification and minimization strategies through its corrective powers, based on Article 58(2) GDPR. Lastly, it is worth stressing that heavy administrative fines may be imposed under Article 83 of GDPR in respect of all the infringements of the said GDPR provisions (Hacker, 2018).

The idea that data protection law could address diversity biases of AI applications in the labor market has sometimes been challenged and accompanied by a call for caution. In particular, Frederik J. Zuiderveen Borgesius is critical for three reasons. He first emphasizes the existing compliance and enforcement deficit specific to the GDPR. Second, the scholar explains that, with data protection law only applying to data processes, the regulation of AI systems could sometimes fall outside its scope. Third, no matter the fundamental principles and rights the GDPR enshrines, AI systems often turn into black boxes beyond human control. (Zuiderveen Borgesius, 2020).

Funded by the European Union

email info@biasproject.eu

3.2.7 Addressing diversity biases in AI applications in the labor market: Relevant EU legislative proposals

The previous section discussed data protection law and its potential to address diversity biases of AI applications in the labor market, particularly regarding selection and recruitment practices. This section touches upon the proposal for a Regulation laying down harmonized rules on artificial intelligence (from now on, the AI Act) and the proposal for a Directive on improving the working conditions in platform work (hereinafter Platform Work Directive). Given their possible adoption during the implementation of the BIAS project, the Consortium considered it necessary to briefly examine their possible regulatory impact on the identification and mitigation of diversity biases under scrutiny.

In 2021, the European Commission presented the AI Act, which attempts to lay down minimum rules to ensure the development, launch on the market, and use of safe and legal systems, to respect the principle of legal certainty, to comply with the fundamental rights and safety requirements already in force, and facilitate the single market.[4] For this purpose, it classifies AI according to its design and the related level of risk and divides it into four categories, namely unacceptable risks (Title II), high risks (Title III), limited risks (Title IV), and minimal risks (Title IX). As the level of risk increases, stricter legal rules apply (Veale & Zuiderveen Borgesius, 2021). As far as the deployment of AI applications in the labor market is concerned, Recital No. 36 of the AI Act explicitly provides that:

> "AI systems used in employment, workers management and access to self-employment, notably for the recruitment and selection of persons, for making decisions on promotion and termination and for task allocation, monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may appreciably impact future career prospects and livelihoods of these persons".

As such, any AI application in the labor market is expected to comply with several requirements and, more precisely, the establishment, implementation, documentation, and maintenance of a risk management system (Article 9), the respect for quality criteria of data training, validation, and testing, such as relevance, completeness, and diversity (Article 10), the maintenance of technical documentation proving the lawfulness of the AI system (Article 11), the automatic recording of events while the AI system is operating (Article 12), the respect for transparency (Article 13), the guarantee of human oversight (Article 14), the conformity with a design that is accurate, robust, and cyber-secured (Article 15). Compliance with all these requirements leads to the performance of conformity assessment to provide evidence. In this regard, it is important to stress that the BIAS Consortium already plans to run the assessment list of trustworthy AI (also known as the ALTAI) as part of WP3, thereby self-evaluating whether the design of the Debiaser complies with human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and

[4] At the time of internally reviewing this Deliverable, the second trialogue should start taking place. For more detailed information about the legislative train schedule, see: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence

Funded by the European Union

environmental well-being, and accountability. In any case, given the recent nature of this proposal, it is still unknown whether the AI Act could be effective in identifying and mitigating diversity biases of AI applications in the labor market, nor much literature has been produced for this purpose. However, Antonio Aloisi is already critical of its preventive nature, have de-regulatory effects on the current sectoral and national legislation, which sets a higher standard of protection (Aloisi, 2024).

In 2021, the European Commission also presented the Platform Work Directive, the aim being to improve the working conditions of people engaging in digital labor platforms.[5] For this purpose, it promotes transparency, fairness, and accountability of the AI system underpinning the operation of platforms, with further attention to the processing of personal data (Chapters III and IV) and providing the workers with legal remedies (Chapter V). For instance, Article 6 requires digital labor platforms to inform workers about the use and functioning of any AI application used to monitor, supervise, or asses their work performance. Also, based on Recital 37, "persons performing platform work should have the right to obtain an explanation from the digital labor platform for a decision, the lack of decision or a set of decisions taken or supported by automated systems that significantly affect their working conditions […] Where such decisions infringe those persons' rights, such as labor rights or the right to non-discrimination, the digital labor platform should rectify such decisions without delay or, where that is not possible, provide adequate compensation." Although the Platform Work Directive does not explicitly make it clear, it appears that all these provisions could and aim to respond to the diversity biases of digital platforms. Indeed, the use of algorithmic management has already proved to lead to direct and indirect discrimination due to its underlying diversity biases, as well as cause significant harm to digital platform workers who, incidentally, are primarily people already facing barriers in access to the labor market, like migrants, women, and people with disabilities (EU-OSHA, 2023). In any case, as was the case with the AI Act, it remains to be seen whether the Platform Work Directive will be able to reach its goals effectively, and a critical body of literature is gradually growing (De Leo & Grossi, 2023) (Aloisi & Potocka-Sionek, 2022).

3.2.8 Addressing diversity biases in AI applications in the labor market: Anti-discrimination law

This section seeks to understand whether anti-discrimination law could capture diversity biases of AI applications for selection and recruitment purposes and provide effective legal remedies. Briefly, although diversity biases arising from the deployment of AI applications fall within the scope of anti-discrimination law, they pose some challenges to its effectiveness because AI applications are likely to discriminate in a way that is different from human beings.

A comprehensive explanation of anti-discrimination law falls outside the scope of this deliverable. Nonetheless, it is still worth getting back to basics to better understand the ambitions and limitations of anti-discrimination law in addressing diversity biases of AI applications in the labor market. In a nutshell, the EU anti-discrimination law is composed of primary and secondary sources. As far as primary sources are concerned, Article 21 of the European Charter of Fundamental Rights provides that "[a]ny discrimination based on any

[5] This means that the the Platform Work Directive does not explicitly cover the recruitment and selection process.

Funded by the European Union

ground such as sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited" (See further: Kilpatrick & Eklund, 2021). At the level of secondary sources, a few directives set a general framework, which lists certain legal grounds and prohibits discrimination against them.[6] In this regard, anti-discrimination law explicitly distinguishes between direct discrimination and indirect discrimination.

- **Direct discrimination** refers to situations in which someone is treated less favorably than another is, has been or would be treated in a comparable situation, due to their personal characteristics. With some exceptions (Adams-Prassl et al., 2023), it is commonly assumed that direct discrimination seldom arises from the design and deployment of AI applications (Aloisi, 2024)(Hacker, 2018). In the labor market, for example, it is unlikely for AI applications to explicitly use gender, race, or another legally protected ground and to attribute lower scores to it, while matching a job application with a job candidate.
- **Indirect discrimination** covers all those situations where an apparently neutral provision, criterion, or practice put a person or a group of people at a particular disadvantage, in comparison with other people, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary. In the case of AI applications in the labor market, this means, for instance, that an AI-driven selection and recruitment process will cause indirect discrimination, if it unjustifiably rejects job applications from a disproportionate number of people due to their personal characteristics. Overall, indirect discrimination arising from the design and deployment of AI applications is considered to be more common to arise because this technology mostly relies on neutral criteria and practices (Aloisi, 2024) (Hacker, 2018).

Outside the current scope of anti-discrimination law and its definitions, most literature on AI-driven discrimination refers to the concept of 'proxy discrimination' in order to cover all those forms of discrimination arising from the correlation with protected grounds (Xenidis, 2020) (Kim, 2019)(Kleinberg et al., 2018). In other words, it is likely for proxy discrimination to occur whenever the AI application uses neutral information as a stand-in (*i.e.*, a proxy) for a prohibited ground. This means, for instance, that the address of the job applicant is asked to infer their race or ethnicity (Bertrand & Mullainathan, 2004). According to Daniel Schwarcz and Anya E. R. Prince, proxy discrimination is far from a novelty, and employees have traditionally resorted to it to thwart anti-discrimination law. At the same time, the scholars acknowledge that, while HR practitioners seldom discriminate on the grounds of proxy information unintentionally, it is likely for an AI application to do so because:

> "The inherent tendency of AIs to engage in proxy discrimination when they are deprived of directly predictive traits follows

[6] The main non-discrimination directives, that are relevant to the labor market are the following: the Racial Equality Directive (2000) prohibiting discrimination on the basis or racial or ethnic origin in many contexts; the Employment Equality Directive (2000) prohibiting discrimination based on religion or belief, disability, age, or sexual orientation in the employment context; and the Recast Directive (2006) prohibiting discrimination based on gender in employment and occupation. For a general overview of EU anti-discrimination law, see: (FRA, 2018)

> inextricably from their structure. Predictive AIs are programmed to locate correlations between input data and target variables of interest. But unlike traditional statistical models, AIs do not accomplish this by relying on a human's starting intuition about causal explanations for statistical linkages between input data and the target variable. Instead, AIs use training data to discover on their own what characteristics can be used to predict the target variable. Although this process completely ignores causation, it results in AIs inevitably "seeking out" proxies for directly predictive characteristics when data on these characteristics is not made available to the AI due to legal prohibitions. Simply denying AIs access to the most intuitive proxies for directly predictive variables does little to thwart this process; instead it simply causes AIs to produce models that rely on less intuitive proxies" (Daniel Schwarcz & Anya E.R. Prince, 2020, p. 1264-1265).

With this in mind, it appears that proxy discrimination arising from AI applications places some significant limitations on the scope of anti-discrimination law, considering that a person is discriminated against based on a piece of information that is intertwined but other than the ones legally qualified as protected grounds. While the addition of new protected grounds could be a *prima facie* option, it is more commonly held that the legal category of discrimination by association *(i.e.*, where the person who is discriminated does not have a protected group but is associated with a person having it) could cover all those situations (Aloisi, 2024)(Xenidis, 2020). This could be the case, for instance, if the AI application processes behavioral data to determine the productivity of a job applicant and rejects the job application of a person with past experiences of depression in their family and recurring sick leaves in previous job positions.

More broadly, Sandra Wachter highlights that discrimination in AI applications cannot only be traced back to traditional asymmetries of power but should also include social groups that do not currently enjoy legal protection and share some personal characteristics that human beings can easily identify or not (*i.e.*, human-comprehensible or incomprehensible characteristics). This is respectively the case of, on the one hand, single parents and homeless people and, on the other hand, people with similar web history or mouse movements (Wachter, 2022). Again, it appears that the solution to this limitation lies in the change for the legislator to more broadly or differently define the current list of protected grounds, with Sandra Wachter also stressing the importance of the AI application to comply with certain requirements, like stability, transparency, empirical coherence, and ethical and normative acceptability (Wachter, 2022).

Another limitation of anti-discrimination law in addressing diversity biases of AI applications is its failure to consider intersectionality. Briefly, intersectional discrimination covers all those situations where a person faces a *sui generis* form of cumulative discrimination based on multiple personal characteristics interacting simultaneously (Crenshaw, 1989; Fosch-Villaronga et al., 2021). Whereas intersectional discrimination is already pervasive in the analog world, deploying AI applications will likely amplify and give rise to new forms. Indeed, AI applications are expected to amplify and give rise to new forms of intersectional discrimination through datasets structured along intersecting axes of social inequality. However, as anti-discrimination law explicitly addresses intersectional discrimination, and the EU case law has traditionally never responded to it, deploying AI applications inevitably leads to this failure (Xenidis, 2020).

From the litigation angle, Antonio Aloisi argues that it is potentially complicated for the worker to prove direct discrimination, especially in terms of the twofold nexus of causality between the conduct and the harm suffered, as well as between the conduct and the protected ground (Aloisi, 2024). In this regard, it is worth emphasizing that AI applications operate at speeds, scales, and levels of complexity that go beyond human understanding and act upon data that are often not historically protected. This means, on the one hand, that potential victims are not aware of the ongoing discrimination and fail to raise a claim. On the other hand, it is complicated to present evidence due to the lack of transparency and explainability of AI applications (Wachter et al., 2021). Besides, both EU and national case law have never adopted a consistent approach to evaluating *prima facie* discrimination, meaning that anti-discrimination law cannot help the AI (and, in the instant case, the HR and the job seekers') community transpose clear and constant, legal requirements into the design of new technology aimed at the identification and mitigation of diversity biases (Wachter et al., 2021) (Sánchez-Monedero et al., 2020).

3.2.9 Addressing diversity biases in AI applications in the labor market: Alternative measures to the law

According to Philip Hacker, bias identification strategies generally draw on statistical methods that could demonstrate that a person or a group of people have been treated differently. Instead, bias mitigation strategies could be divided into three groups, namely pre-processing techniques re-balancing the dataset, in-processing techniques incorporating changes into the objective function or imposing a fairness constraint, and post-processing techniques re-assigning the labels to a fairer state (Colmenarejo et al., 2022).[7] As previously mentioned, a growing body of technical literature currently aims at the creation of fairness metrics. With NTNU and BFH examining this specific literature in a separate working paper,[8] this subsection briefly focuses on some exemplifying, alternative and/or ancillary measures to the law that could contribute to the identification and mitigation of diversity bias and might somewhat relate to the design of the Debiaser.

It is necessary to process and know their characteristics to demonstrate whether an AI application discriminates against a job applicant. Also, it will be unlikely for an employer to account for job applicants different from the ones they have already hired if the AI application cannot detect diverse personal characteristics and only relies on previous selection and recruitment procedures (Barocas & Selbst, 2016). In the European Union, though, Article 9 of the GDPR prohibits the processing of special categories of data, including race, ethnicity, health, genetics, biometrics, religion, political and philosophical beliefs, trade union memberships, and sexual orientation. This means that, at present, the GDPR does not allow the processing of special categories of data for debiasing (Van Bekkum & Zuiderveen Borgesius, 2023). On this

[7] For the sake of completeness, it worth highlighting that other authors suggest different classification systems, such as (Hacker, 2018)

[8] See footnote no. 1.

point, Betsy A. Williams et al. argue that this legal constraint ignores or hides – rather than addresses – discrimination, and the ongoing AI-driven discrimination currently proves it (Williams et al., 2018). Accordingly, the fairness, accountability, and transparency community (from now on, the FAccT community) calls for greater collection of special categories of data to facilitate the identification of diversity biases and eventually file a lawsuit (Van Bekkum & Zuiderveen Borgesius, 2023) (Schwarcz & Prince, 2020) (Gillis & Spiess, 2019) (Kleinberg et al., 2018) (Williams et al., 2018). Some policymakers have already adopted or are about to introduce an exception to the prohibition on processing special data categories for non-discrimination (Van Bekkum & Zuiderveen Borgesius, 2023). In particular, Article 10(5) of the AI Act grants this permission, though limiting it to the deployment of high-risk AI applications and making it contingent on compliance with appropriate safeguards.[9] In any case, Sandra Wachter et al. clarify that design with special categories of data cannot solve the problem of diversity bias on its own (Wachter et al., 2021), meaning that other identification and mitigation strategies should work in synergy.

This is the case, for example, of the performance of an audit. In broad terms, an audit attempts to understand how the AI application plays out in practice, especially to evaluate whether it negatively affects some legal, ethical, and social interests or values (Koshiyama et al., 2021) (Raub McKenzie, 2018) (Sandvig et al., 2014). In describing how to perform an audit, some authors emphasize the negative outcome arising from deploying the technology (Brown et al., 2021), while others focus on formal compliance with applicable norms (Mökander et al., 2021). Also, Matti Minkkinen et al. suggest the introduction of continuous auditing, which could keep pace with the design and deployment of the AI application and could continuously provide up-to-date information (Minkkinen et al., 2022).

Another strategy for bias identification and mitigation relies on the design process of the technology, on the assumption that design can have a significant impact on our conduct. For example, when the traffic light turns red, a pedestrian, a rider, and a driver are expected to stop because of their normative function. As far as the design of more advanced technologies - including AI applications - is concerned, it is likewise commonly held that it transposes ethical, moral, and legal norms into its software and hardware, thereby nudging our conduct (van den Berg & Leenes, 2013). With some nuances, this theory applies the labels of 'techno-regulation' or 'legal by design' (Hildebrandt, 2020), as well as value-sensitive design and values in design (Aizenberg & van den Hoven, 2020). Although few related studies address diversity bias, Judith Simon et al. explicitly supports value-sensitive design because it could go beyond the mere adoption of technical measures aimed at the prevention or avoidance of discrimination while instead prioritizing the social, political, and economic context where diversity bias arises (Simon et al., 2020). In this regard, it is worth mentioning that, at the time of writing, ULEID and BFH are drafting an article to be submitted to the Third Workshop on Bias and Fairness in AI at the

[9] In terms of technological solutions to the issue of data privacy, computer scientists are actively working on privacy-preserving data analytic methods that rely on the notion of differential privacy in building algorithms. Here, data are randomized during the collection process, which leads to "learning nothing about an individual while learning useful information about the population (Tambe et al., 2019)

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (Turin, 18-22 September 2023).

The idea that technical tweaks are more limited in scope and should be considered short-term solutions is also shared within the FAccT community, which pushes for structural reform both on a regulatory level and within the private sector. While some authors argue for the establishment of a regulatory body that could oversee the development of new technologies before their launch on the market, a growing body of literature highlights the lack of diversity in the tech industry and its following homogeneity in worldviews (Raub McKenzie, 2018) (Williams et al., 2018). Because each AI application embeds human worldviews in code and could therefore include human bias too, most scholars welcome a participatory or co-creation approach (Birhane et al., 2022)(Hossain & Ahmed, 2021)(Zytko et al., 2022)(Mucha et al., 2020). At the heart of participatory and co-creation approaches lies the potential to consider the needs and views of social communities that have traditionally been marginalized, thereby leading to a more diverse and inclusive technological design and affordance (Poulsen et al., 2020). In this regard, it is important to stress that a co-creation methodology is one of the founding pillars of the BIAS project.[10]

Lastly, it is worth mentioning transparency. Sometimes transparency is generalized and equated with a call for disclosure of information about the possible impact AI applications have on a person or a social group, especially when legally protected due to their personal characteristics (Schwarcz & Prince, 2020). Instead, Jon Kleinberg et al. suggest regulating in detail the process through which the AI application is designed while including record-keeping requirements. For the authors, "[t]he opacity of the algorithm does not prevent us from scrutinizing its construction or experimenting with its behavior—two activities that are impossible with humans [...who] disassemble, obfuscate, and lie" (Kleinberg et al., 2018, p. 118). Accordingly, transparency-driven regulation should target every human choice in the development of the AI application and, more specifically, each decision about 1) what outcome to predict, 2) what inputs to make available to the technology for consideration, and 3) the training procedure itself (Kleinberg et al., 2018).

[10] On the co-creation methodology, see Deliverables 2.2 and 2.4 covering the co-creation methodologies and findings.

4 The national labs

The national lab is a pool of diverse stakeholders, including employees, employers, HR practitioners, AI developers, policymakers, trade union representatives, civil-based society organizations, scholars, and other people whose professional expertise and/or experience could contribute to implementing the BIAS project. With the creation of a national lab in each project country, the BIAS Consortium aims to adopt a co-creation approach that could inform the design of the Debiaser in a way that is as inclusive and diverse as possible. Additionally, the national labs allow for facilitating networking amongst people working on AI and diversity in the labor market.[11]

4.1 The creation process

To create the national labs, ULEID first drafted an information sheet and an informed consent form (hereinafter the national labs consent form) in December 2022. This document briefly explains to possible stakeholders the BIAS project and, more precisely, the purpose and implementation of the national labs, as well as to obtain their consent to invite them later to participate in project activities.

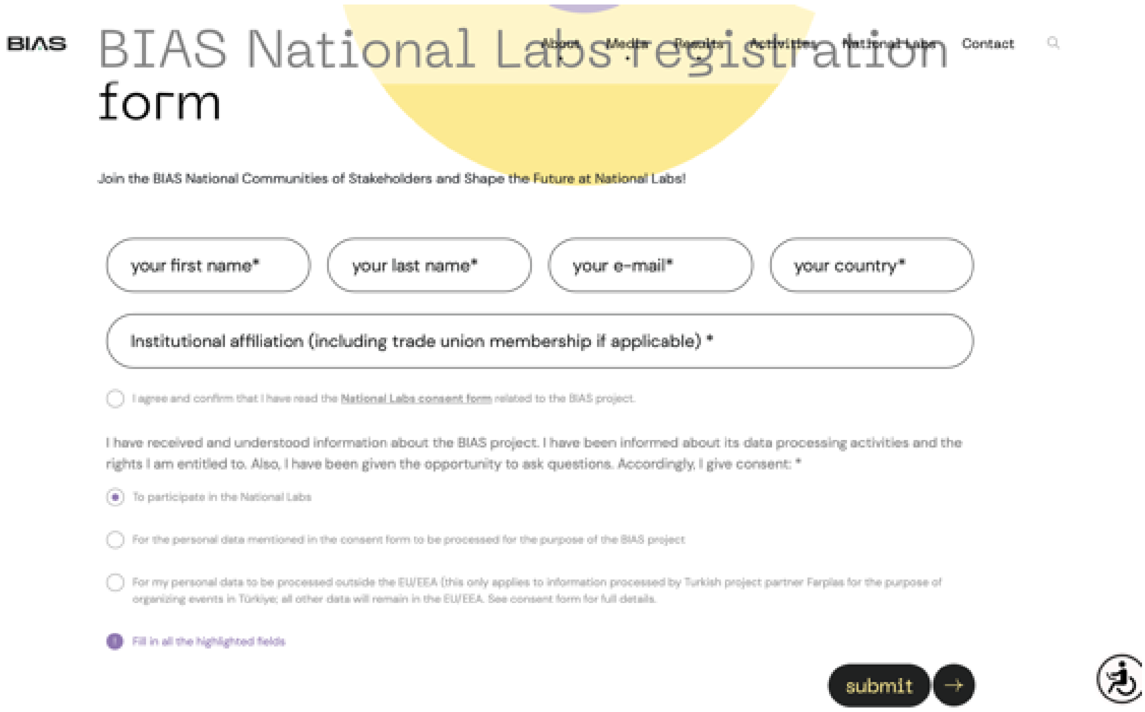

Figure 1 National Labs registration form as seen in the project's website

After being internally revised by NTNU and SVEN, this draft was submitted to the Norwegian Data Protection Authority for data protection clearance. The national labs' consent form collects some personal data, namely the name, the e-mail address, the institutional affiliation (which, is not mandatory to report), and the respondent's home country. The Norwegian Data Protection

[11] On more information about the networking opportunities arising from the creation of the national labs and the co-creation approach, see: D2.2 and D2.3 on co-creation methodologies and findings.

Authority green-lighted the National Labs consent form in January 2023. ULEID, therefore, started to discuss the layout and the dissemination strategy, especially with LOBA and NTNU. Also, ULEID decided to add the creation of an international lab, which could include all those stakeholders that, though coming from European countries other than the ones composing the BIAS Consortium, could still contribute to the implementation of the BIAS project because of the professional experience and/or expertise. Since it was technically complicated to transpose the national labs consent form into an online registry where each stakeholder could put their name, date, and signature manually, the registration process first involved the downloading of the document from the BIAS website and the following submission to the BIAS project e-mail (*i.e.*, info@biasproject.eu).

This registration process was developed and completed in March 2023, when the BIAS website was launched. However, some Consortium partners immediately voiced concern about its effectiveness, in the sense that they feared that this two-fold process could discourage some stakeholders from joining the national labs. Accordingly, ULEID, LOBA, and NTNU found an *interim* solution involving the creation of an online form of Wufoo, namely a web application for creating online forms, surveys, and event registrations that is user- and data-protection-friendly.

Until May 2023, anyone who was interested in the BIAS project could join the national labs by going on the BIAS website, clicking on the green button 'Apply Now', and filling in the online form on Wufoo. The filled-in form was directly sent to the BIAS e-mail address, which is accessible to LOBA (as the WP7 leader on communication, dissemination, and exploitation), ULEID (as the WP2 on the stakeholder involvement, needs assessment, and co-creation), and NTNU (as the project coordinator). In June 2023, LOBA finalized the registration process of the national labs on the project website. This registration process will be available until the end of the BIAS project (*i.e.*, October 2025), even though ULEID already aims at reaching the threshold of 100 stakeholders per partner by M15 (*i.e.*, January 2023) when the final report on co-creation methodologies and findings is due. Annex II includes a copy of the national labs information forms.

4.2 The dissemination and communication process

At the time of writing, each Consortium partner is getting in contact with national and international stakeholders to create the national labs based on the dissemination strategy LOBA developed, together with ULEID. In a nutshell, the dissemination strategy includes the following activities:

1. A dedicated page on the BIAS website that explains the reason, the purpose, and the functioning of the national labs, as well as how to apply to join them.
2. Four templates of e-mails Consortium partners can use to get in contact with possible stakeholders, depending on whether they already know them or ask for active or passive involvement in the project implementation.
3. A social media campaign, where LOBA and each Consortium partner advertise the creation of the national labs and invite people to join them. All the posts have been also translated into all the national languages of the BIAS Consortium.

4. An infographic designed to present the goals and features of the National Labs in a visual way.

More detailed information about the dissemination and communication strategy for the creation of the national labs will be included in D7.2 on the revised dissemination, exploitation, and communication plan, which is due in M24 (*i.e.*, October 2024).

4.3 The present and future of the national labs

ULEID plans to reach the threshold of 100 stakeholders per partner by M15 (*i.e.*, January 2023) when the final report D2.3 is due. However, national and international stakeholders will be able to join the national labs until the end of the project. Even though the co-creation process will come to an end in M42, the national labs could generally channel other project tasks, including ethnographic research (WP4) and capacity building and raising awareness (WP5), as well as be the target of most communication and dissemination activities (WP7).

For this purpose, the BIAS Consortium is inviting possible stakeholders to participate in the BIAS project and its different activities, with the first iteration of co-creation workshops in June and July 2023 being the first occasion. How many and precisely which activities each stakeholder takes part in is up to the stakeholders. Accordingly, it is possible to divide the members of the national labs into two main groups. On the one hand, some stakeholders are interested in the BIAS project but would rather keep a low profile and passively receive some news about its implementation. On the other hand, other stakeholders are willing to be proactively involved, according to their professional expertise, experience, and interest. This has already been the case with the respondents to the expert interviews (T2.3).

Funded by the European Union

5 The mapping exercise of existing AI applications

The mapping exercise sought to scope the current design and/or deployment of AI applications for HR purposes. The mapping exercise therefore contributes to the performance of the needs analysis of WP2, by gaining consolidated knowledge about the technologies already available on the market and the specific gaps in the identification and mitigation of diversity biases that the BIAS project is expected to fill.

5.1 Setting the scene of the mapping exercise

In December 2022, ULEID drafted an online form so that Consortium partners could map current AI applications for HR management in their own country by simply seeking information via the Internet. The reason behind this methodological choice lies in the change of perspective. With the literature review collecting academic expertise, the expert interviews targeting AI developers and AI practitioners, and the survey covering the personal experiences of workers, this online search could help the BIAS Consortium put itself in the place of 'the average person' trying to understand how AI applications in the labor market work. Indeed, none of the BIAS researchers doing the mapping exercise had previous experience with the technologies they identified and reported to ULEID.

The overall structure of the online form was composed of three sections. The first section aimed to identify the current AI applications by reporting their names and the websites where to collect information. The second section attempted to contextualize each AI application and its specificities. For this purpose, Consortium partners were asked to clarify where, why, and how the AI application was designed and/or deployed. The third section shifted the focus toward diversity issues, especially regarding data processing operations, mitigation measures, and fairness perceptions.

After being internally revised, ULEID shared the online form amongst all those Consortium partners involved in T2.1, namely NTNU, BFH, HI, SVEN, DIGI, and FARPL. The aim was to target circa five AI applications per country. ULEID set the timeframe for the mapping exercise between January and February 2023 to conduct preliminary data analysis in March 2023. Based on the resulting research outputs, ULEID ultimately did some additional research to double-check specific information, examine certain issues more in detail, and create some synergies with other WP2 tasks. The template of the mapping exercise is included in Annex III of this deliverable.

5.2 Research findings

As a result of the mapping exercise, ULEID examined thirty-five AI applications for HR management that have already been launched on the market. Although the geographical scope originally lay in the home countries of the Consortium partners involved in T2.1, it was later broadened to cover Europe. Indeed, it soon appeared that the design and/or deployment of AI applications for HR management usually go beyond national boundaries, meaning that they are typically manufactured, sold, and/or used in more than one country, as well as in countries other than Estonia, Iceland, Italy, The Netherlands, Norway, Switzerland, and Türkiye.

Funded by the European Union

email info@biasproject.eu

A summary and analysis of the research findings follow, divided into two subsections according to the specificities and diversity considerations of the AI applications. Annex IV lists the AI applications that have been mapped, while indicating their source.

5.2.1 Specificities of the AI applications for HR management

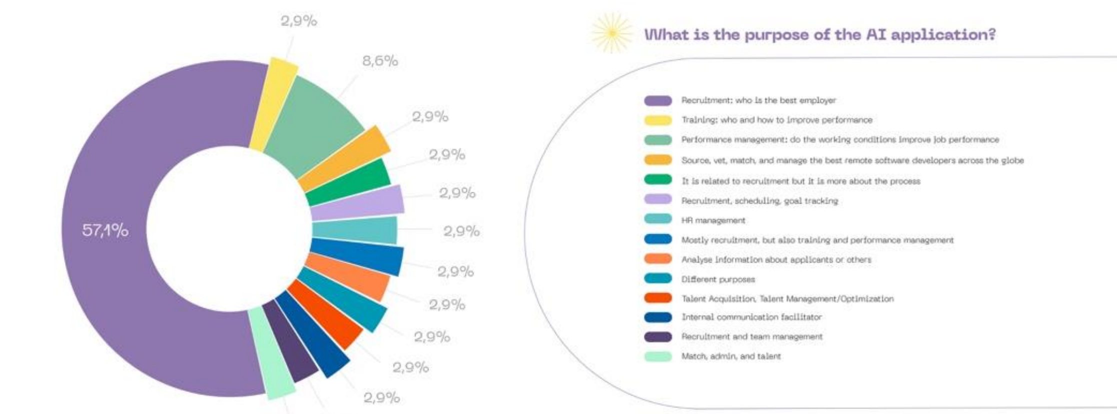

Figure 2: Mapping the purpose(s) of the AI applications.

There are two types of AI applications currently being adopted in the labor market. On the one hand, some AI applications primarily aim at facilitating and/or replacing HR practitioners in the implementation of specific tasks, such as selection and recruitment, time management, and performance assessment. On the other hand, other AI applications attempt to identify and mitigate diversity biases that negatively affect HR management, be it manually or automatically done. Whilst it was interesting for ULEID to understand whether and how the former category of AI applications could receive the benefit of the development of the Debiaser, the latter could help the BIAS Consortium distinguish its innovative technology from hi-tech devices already existing on the market.

No matter whether the AI application is designed to perform specific HR tasks or address diversity biases in HR management, there is some evidence to suggest that its deployment mostly takes place in the private sector and, more precisely, in big size but diverse companies (*e.g.*, Booking.com, Heineken, Whirlpool, and KLM). As far as the public sector is concerned, it is worth observing that the deployment targets public administration offices (*e.g.,* Stadt Zürich), hospitals (*e.g.,* Kantonspital St.Gallen), and universities (*e.g.*, Leiden University). In any case, it is normally advertised that the deployment of AI applications for HR management is likely to improve the efficiency of the HR team, especially in terms of time, money, and effort.

Whilst it is clear that all the AI applications seeking to identify and mitigate diversity biases in HR management aim at ensuring equality, the remaining AI applications vary in their purpose(s). More frequently, AI applications are manufactured, sold, and used to facilitate selection and recruitment procedures. This means that the highest number of technologies attempt to attract job applicants, as well as file and screen applications, in order to find the best match. Another common purpose is for the AI application to make the HR management of the individual worker, as well as the working team, agile and effective. Instead, no AI application designed to control employees due to disciplinary motives or productivity was found. The latter research output

Funded by the European Union

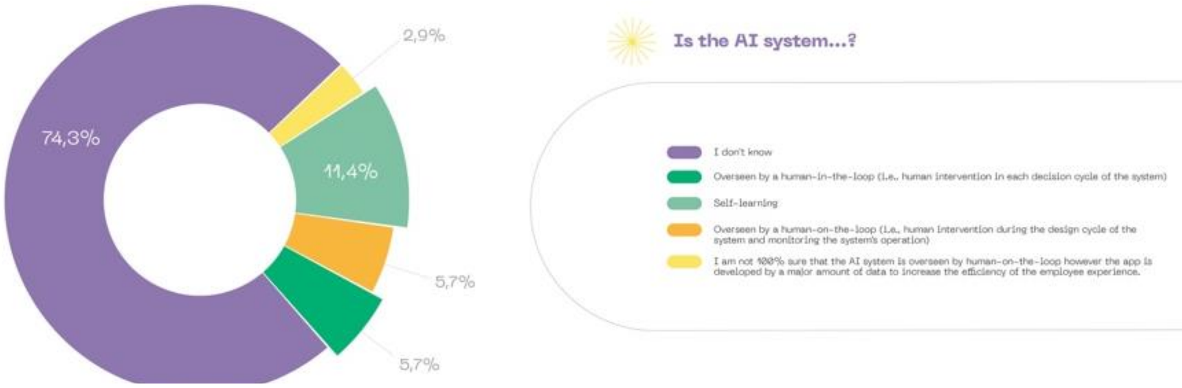

Figure 3 Mapping the design of the AI applications.

could be read together with the ones arising from the expert interviews and reported in section 6 of this deliverable, where many respondents cautioned against the deployment of AI applications for this purpose.

Be they designed to perform specific HR tasks or address diversity biases in HR management, most AI applications are designed to help HR practitioners perform their tasks, especially in view of a decision to take about a candidate or a worker. Alternatively, some AI applications take the

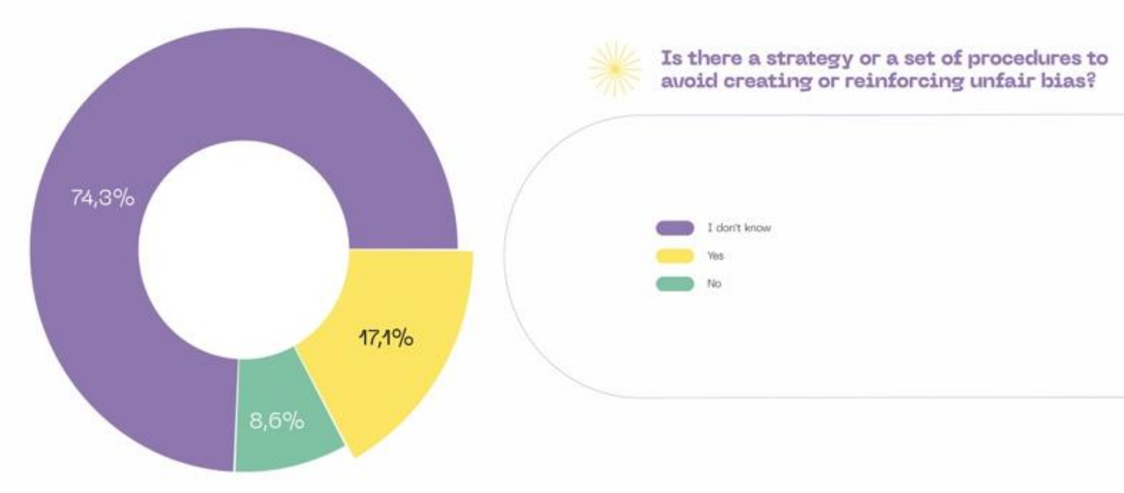

Figure 4: Mapping any bias mitigation strategy of the AI applications.

form of a chatbot and interact with candidates or workers for different reasons (*e.g.*, solving doubts or setting a meeting). To the knowledge of the BIAS Consortium, there is no AI application that is designed to make decisions with a degree of autonomy.

All the information available on the internet did not allow ULEID and the other Consortium partners to understand whether the job applicant or the employer interacting with an AI application was informed about its possible deployment. Out of the 35 AI applications ULEID analyzed, it appears that only one explicitly warns the user about the engagement with an AI application through the previous submission of an informed consent form. Additionally, it was

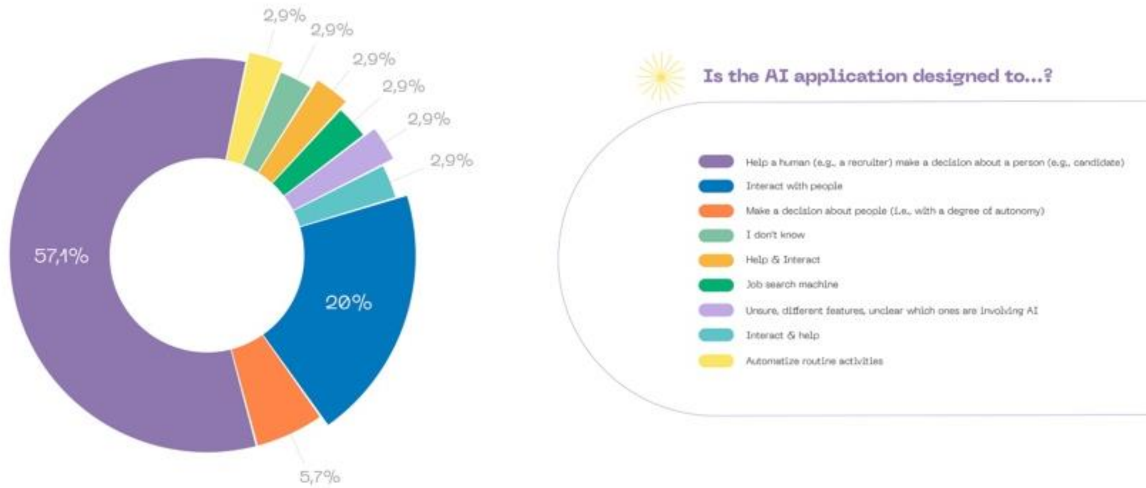

Figure 5: Mapping the autonomous nature of the AI applications.

not explicitly clarified whether the AI system is self-learning or is overseen by a human-in-the-loop *(i.e.*, there is human intervention in each decision cycle of the system), a human-on-the-loop (*i.e.*, there is human intervention during the design cycle of the system and human monitoring of the system's operations), or a human-in command (*i.e.*, when the human being can decide when and how to use the AI application in any particular situation). At the time of writing, the absence of information is also one of the *interim* results of the survey described in Section 7 of this deliverable.

5.2.2 Diversity considerations in AI applications for HR management

Overall, this mapping exercise demonstrates that most AI applications for HR management share a lack of transparency in how diversity considerations are presented online. Instead, several websites advertising these technologies generally emphasize compliance with the GDPR and other national provisions ensuring data privacy and security.

More precisely, ULEID and the other Consortium partners could not examine the datasets used to design the AI applications (especially in terms of their representativeness, completeness, and recentness), the personal data processed by the AI applications process, and the adoption of a mechanism allowing for the flagging of issues related to bias, discrimination, or the poor performance of the AI application. Besides, it seems that no AI application not designed in a way that people with special needs or disabilities could use.

When examining AI applications doing HR tasks, it is not common to find a strategy or a set of procedures to avoid creating or reinforcing diversity biases. However, when these identification and mitigation measures are in place, they frequently range over the performance of an impact assessment, adopting a code of conduct, providing diversity training to the staff, and using other AI applications similar to the Debiaser. This research output is somewhat in line with the findings of the expert interviews described in Section 7 of this deliverable, as well as some legal and alternative measures discussed in the literature review of Section 3.

In conclusion, the lack of transparency made it impossible for the Consortium partners to determine whether the existing AI applications were fair. At most, it is possible to observe that

Funded by the European Union
email
info@biasproject.eu
website
www.biasproject.eu

many technologies designed to identify and mitigate diversity biases seem to embody an ideal of equal treatment between individuals, in the sense that they seem to address diversity biases without positively discriminating individuals.

Funded by the European Union

6 The expert interviews

The BIAS Consortium examined the professional experiences and personal attitudes of HR executives and AI developers to gain better knowledge about the diversity biases of AI applications in the labor market. Qualitative research and, more precisely, semi-structured interviews were expected to understand better the ambitions and limitations that AI applications in the labor market could have, especially regarding bias creation, identification, and mitigation. The interviews concentrated on selection, recruitment procedures, and HR management. The latter focus provided a bridge between this task and the performance of ethnographic fieldwork in WP4. Overall, the expert interviews demonstrate a positive attitude towards AI-driven practices for selection and recruitment, with AI-driven management being a more sensitive issue. It appears that the risk of bias is high, especially when reproducing human ones and relating to gender. Some legal and ethical requirements (*e.g.*, transparency and human oversight) and mitigation measures (*e.g.*, diversity quotas) are identified.

6.1 Methodology

6.1.1 Drafting the survey

As the task leader of this project activity, HI first drafted two semi-structured interviews that could respectively target HR practitioners and AI developers in November 2022. The aim was to understand their professional experiences and personal attitudes towards diversity biases and AI applications in the labor market. NTNU, ULEID, and SVEN then revised the draft, which became structured as follows after a round-table discussion. In both documents, the interview begins with a general question about the type of organization the respondent works in and finishes with identifying topics of common interest to be addressed in future project activities, including the national labs (T2.1) and capacity building (WP5). All the remaining questions are divided into two groups according to their connection with selection and recruitment procedures or HR management.

On the one hand, AI developers are primarily asked to discuss the creation, reproduction, and mitigation of diversity biases, the functioning of data processing operations, and their familiarity with AI and ethics. On the other hand, the questions for HR executives primarily focus on their professional opinions, especially on the potential impact of AI-driven practices for recruitment and management purposes. A final question covers the willingness to join training on diversity and HR in the labor market and, if so, some specific suggestions. The templates of the semi-structured interviews are included in Annex V.

In December 2022, NTNU shared the final version of the semi-structured interviews with the Norwegian Data Protection Authority to get the green light a month later. Precisely, the Norwegian Data Protection Authority recognized that the way the expert interviews had been structured did not process personal data, with all the questions but one focusing on the professional experience and expertise of the respondent. Indeed, gender is the only information collected for this deliverable, in order to prove the respect of gender balance to the European Commission.

Funded by the European Union

6.1.2 Recruiting the respondents

To gather a relevant sample of respondents, all the BIAS partners but LOBA and Crowdhelix – not involved in this project activity - started identifying candidates from various organizations in each respective country in December 2022. In doing so, the BIAS partners sought AI developers and HR executives able to discuss the topic of diversity biases of AI applications in the labor market within their professional expertise and experience. Therefore, they examined each candidate's professional profile and approachability for interviews. While some candidates were found through the existing network of the Consortium, others were identified through random targeting. As a result, each partner invited the relevant people to the expert interview study via email and tried to have ten people agree to participate while respecting gender balance as much as possible. Out of the 71 interviews, 17 women and 18 men composed the AI community, while 28 women and 8 men constituted the HR community. To gather a relevant sample of respondents, all the BIAS partners but LOBA and Crowdhelix – not involved in this project activity - started identifying candidates from various organizations in each respective country in December 2022. In doing so, the BIAS partners sought AI developers and HR executives able to discuss the topic of diversity biases of AI applications in the labor market within their professional expertise and experience. Therefore, they examined each candidate's professional profile and approachability for interviews. While some candidates were found through the existing network of the Consortium, others were identified through random targeting. As a result, each partner invited the relevant people to the expert interview study via email and tried to have ten people agree to participate while respecting gender balance as much as possible. Out of the 71 interviews, 17 women and 18 men were in the AI community, while 28 women and 8 men were in the HR community.

6.1.3 Conducting the expert interviews

Except for Digitouch - which asked for a deadline extension for internal reasons, all the Consortium partners conducted the semi-structured, expert interviews between December 2022 and April 2023, with one or two researchers present in their performance. In most cases, the physical context was a virtual room – be it on Microsoft Teams or Zoom, depending on the internal IT equipment of the organization, thereby allowing for more flexibility in the interview schedule. From the Consortium partners' experience, this choice did not prevent the creation of a relaxed atmosphere, perhaps due to the increasing normality of online interaction. Also, the Consortium partners allowed the respondents to be interviewed in their native language or English. While some interviews were recorded upon the respondent's consent, others were not. The recorded interviews were temporarily stored in compliance with the applicable data protection law so that the BIAS partners could draft the national report for HI (see below). At the time of writing, all the recordings have been deleted adequately.

6.2 Summarizing the interview results

Interview summaries prepared by Guðbjörg Linda Rafnsdóttir and Dilys Sharona Quartey, University of Iceland

Once the expert interview process was finalized in April 2023, all the involved Consortium partners provided HI with a copy of all the answers they had collected and a national report summarizing the main research outputs. Based on this information, a summary of the responses follows. All replies are the respondents' responsibility and do not reflect the opinions of the BIAS Consortium.

This section is divided into two subsections, which respectively seek to understand the description of biases of AI applications and the deployment of AI applications for recruitment and management purposes. The subsections report the interview questions of the semi-structured template.

6.2.1 AI Developers

The interviewed AI developers worked in public and private organizations or universities, global and national consulting firms, small-medium enterprises, and a start-up. These organizations employ between 10 to 50,000 employees.

Q: When developing an AI solution, biases might occur at different levels: data that is inputted into the model, algorithms performed on the data, and outcomes of the models. Which of these biases do you think are more problematic/difficult to tackle.

The AI experts agreed that biases can easily be present in the dataset. These biases are deep-seated and reflective of historical or societal biases and can be challenging to identify and correct without introducing new biases. Again, if the data is not fed in properly or maintained regularly it becomes problematic; the greatest risk being when the data is obtained or entered very early in the process. It is difficult to identify the bias later. However, it is sometimes possible to correct the bias in the later stages when the algorithm makes available results or suggestions based on the data. The problem here is that the reasoning behind algorithmic decision making is not always understood given the strong computing power resulting from many computer units working in tandem, and the presence of comprehensive data. On one hand, bias dataset was viewed as more difficult to tackle, given that data was core to the AI system. The algorithm learns from the data (which may be skewed or unrepresentative of a certain demographic) and thus was considered the only thing to be looked at when results are biased. There will be deviations in the algorithmic decision-making process if the master data is misleading. Reference was made to the importance of algorithmic output processed from the dataset with regards to interpretation and evaluation.

On another hand, it was pointed out that algorithms could be biased depending on input parameters and that algorithms are usually not designed to be unbiased. Of noteworthy is the fact that data is not always taken and input into the computer. Rather a command such as ‘look at the following areas‘ is given. These data such as conversations or emails are often unstructured. Again, certain speeches are often omitted from formal communication while others are not. The algorithm however reads and interprets these as fact and thus the type of material determines whether bias decreases or increases. Further, algorithms were seen as more problematic to tackle due to the black box (neural network) inside the AI being difficult to decipher. Hence, although data that seems fishy can be scrutinized, the difficulty lies in understanding how and why a conclusion is drawn by the AI algorithms. Additionally, algorithms work within a certain success rate where there is a margin of error. It is important to enter and process comprehensive data, however, even when the data is as comprehensive

Funded by the European Union

as desired, errors can still occur at one point.

It was added that the involvement of people unaware of bias is problematic as it hinders the discovery of bias. It is not uncommon to have prejudices at different levels. The human brain can only be imitated to a limited point; hence bias can be encountered at every stage. Emotions vary from one society to the next, thus algorithms should be able to understand these varieties and differences.

Attention was drawn to the reproduction of diversity bias in recruitment and selection processes by AI applications with regards to gender, sex, race/ethnicity/skin colour/nationality, education, language barrier, age, and background. Due to their visibility, gender and race were considered to be at a higher risk of diversity bias. There are words such as policeman that are inherently biased. The language used in job ads discriminates against older persons and women with more jobs calling for young and dynamic applicants and seemingly addressing males even if there is a statement that encourages females. This can impact the AI training data by introducing bias into the system.

Mitigation strategies pointed to the need for diverse and representative data, although it was admitted that this would be difficult to achieve because it was expensive and time consuming, and societal biases existed. Additionally, observing for bias during the robust testing phase and the recruitment process prior to the AI being ordered to come up with suggestions as well as removing the bias or making the users aware of it was noted. Here the person in charge of the work must figure out which bias needs ridding. De-biasing solutions can be used to rebalance data. The design and implementation of algorithms was said to be more impactful when introducing or removing any bias.

Q: Which kind of bias – if any - are in your opinion more likely to be reproduced/strengthened through AI solutions? Please provide some examples

Experts agreed that technology itself has no inherent bias. There was a consensus that bias lay within the data entered as it reflects biases within the society. It was conveyed that, there was a high likelihood for systematic and input data biases to be reproduced and strengthened through AI. AI systems today have moved from making decisions based on a rule and model that contains statements such as true/false and yes/no to one that does not have rules or 100% yes/no reasoning models to rely on. They are instead based on endless examples of how humans have done things from the past through to the present. The AI goes through vast information accessible for instance on the internet to figure out patterns within the data. The issue with this type of AI system is that it tends to reproduce the exact same biases that exist within this vast data and may even exacerbate certain biases.

One example given was that in the comment systems, people heard the loudest may have the most representation. Another example was that if a generative AI tool is asked to produce photos from the construction industry, it will generate male dominated photos and thus contribute to gender bias. Similarly, it was explained that if all the data on the advancement of academic staff in the position of associate or full professor over the past 60 years is taken, the AI algorithm will read that a professor is more likely to be a man rather than a woman. In this case, it will make suggestions for primarily men to be promoted, thereby, perpetuating gender bias. Reference was made to men who are mostly in positions to program being in privileged positions with privileged blindness and thus unable to see these dangers. Systematic bias was said to arise from behavioural patterns, policies and practices embedded in social structures. For instance, if a hiring algorithm is trained on a historical dataset unrepresentative of a particular racial group, it might learn and reproduce this bias. An added example was the Amazon hiring algorithm that learned to penalise female resumes due to the dominance of males in the tech industry. Diversity bias was stated as being more problematic, particularly at the data and algorithmic levels.

Experts mentioned that other forms of bias reproduced and strengthened through AI solutions include those based on sex, name, age, class, origin/ethnicity/race/nationality/citizenship, disability, religion, pregnancy, level of education, universities attended, salary expectations and digital accessibility (owing cell phones or not). Others pointed out more broadly all kinds of bias that have a social impact, those identified in the data set and reproduced or recreated by the AI algorithm. Some experts also mentioned that different forms of bias are dependent on geographical context such as racial bias in the US or gender bias in Europe as well as from different stages or sectors.

In some cases, biases are very specific; for instance, an AI developer who prefers one setting over another will continue developing the software the same way, e.g., using Microsoft. In reference to automated translations or suggestions for male version of words even when the female versions were correct, it was expressed that in the end it is the outcome that is biased and thus difficult to address. It was emphasized that it is complicated to create, have and/or use data sets that are representative and devoid of human biases, although this is up to AI developers. It also depends on how they are used by HR practitioners; in that they could use the technology to seek out specific information thereby discriminating or favouring certain categories of people. Claims were further made that AI can eliminate and prevent prejudices and biases of HR managers during recruitment by supporting objective evaluation.

Some suggested solutions were carefully defining the purpose of usage and the database as well as the outcome of the application to avoid uncertainty. Additionally, when used effectively and correctly, AI solutions can decrease gender and racial biases within society. It could also support the recruitment, educational processes, performance, and career management stages.

Q: Did you attend any course about ethics in AI within your university curricula or beyond it? which one(s)?

While most of the experts interviewed had not taken any university or certified courses pertaining to ethics in AI, a few had done so at the university level due to personal interests and one had attended seminars/webinars on it. Reason being that although such courses are presently common, this was not the case back in the day. Some experts had taken their education a decade ago when AI was seen as science fiction and the word considered derogatory. Courses at the bachelor's and master's levels are skewed towards technical stuff, however, some experts said they had indirectly encountered this concept during their master's education and sectorial/work experience. An expert who had taken a 1–2-week course of ethic intro to AI courses expressed that the topic was mostly geared towards the influence of AI on the workplace with little on AI bias. Nevertheless, all experts were aware of ethics in AI and expressed interest in furthering knowledge in this area. One expert (a professor) had taken the initiative to apply a gendered lens to their work.

It was stated that the European legislation on data security, private and personal data accounts for the recent debate on ethics. There are clear rules that must be followed and are discussed in class. In most courses that deal with human data and in big conferences, ethics is also discussed. Some experts however questioned whether ethics needs to be taught as a separate course or a specific technique in all or most courses. The importance of ethical questions in academic fields not limited to data and computer science was acknowledged as this is inevitable where AI is concerned. Ethics comes to play where the limits of AI with regards to life, death and health is called to question. This is less so when the effect on people is indirect. A desire for deeper ethical discussions in the tech and engineering sector was expressed.

Recruitment

Q. To what extent is AI (algorithms) technology being produced for use in recruitment and selection?

Experts expressed mixed opinions on the use of AI for recruitment and selection within their various countries and abroad. Some believed AI was used to a large extent, some thought it was moderately used, others thought it was rarely used and one had no idea of its extent. Two experts stated that AI was used for recruitment and selection in their country. Experts explained that in smaller countries where the dataset is rather small AI is not used in advanced ways for recruitment. Others referred to stricter data protection within their country as a reason for its limited use. The US as a country and Europe as a continent were mentioned to have widespread use of AI. There was a consensus in the assertion that larger companies require and are beginning to buy and use AI to manage the vast number of applications they receive to filter application documents and CVs as well as to be as efficient and updated as possible. Some did not find candidate filtration to be very efficient.

An expert pointed out that AI is used in the following ways for higher efficiency: to help sift through hundreds of applications; to identify qualified candidates based on keyword matches, algorithms, and recruitment data; it can respond to candidates via automated messaging; to schedule interviews with those considered qualified. Some experts believed AI is currently not produced for wider usage in recruitment and selection but rather in operational and automation applications. Others made mention of the human factor used in conjunction with AI tools that only play a supporting role. For example, AI is used to support the attraction of female talent through the phrasing of job ads. Nevertheless, there are many opportunities to use technology for good beyond what is done today.

Q. Do you think that the use of AI (algorithms) is in general positive in recruitment and selection?

Most experts thought AI tech was positive and saw it as an opportunity if users were aware of the risks and if there was an element of transparency, human oversight and explicability in recruitment and selection processes. Here the human factor was again seen as essential even

Funded by the European Union

though bias could also emanate from them. Humans are of essence because there are parameters which can confuse the AI. Positivity was further expressed in its efficient; time and cost effectiveness (which allows recruiters to focus on more human centric tasks); its use for initial and subsequent selection/screening and onboarding; its use for simple tasks like searching for keywords as well as its use as a micro-target for recruitment or diversity enhancement. Others saw AI as positive if its purpose is to eliminate unconscious bias, although it must be done well as it could backfire. In particularly the tech and other sectors, a lot of issues related to bias in hiring have been found over time. Over time, young white males with certain experiences and attitudes have been hired. Feeding the AI with this employee data may reinforce this bias. One must be aware of this because all AI learn from information fed into them, which can lead to distortions when blindly followed. It's fine if one is aware of ongoing biases, hiring processes and qualifications. Data driven systems was said to have an advantage in that they can expand on the data and make unanticipated conclusions based on unconventional connections, beyond the scope of humans. It eliminates subjectivity and guarantees accurate and consistent personality analysis. With the right algorithm/model, AI can give quality outputs.

A few experts disagreed with the positivity of AI in this context, stating that it is difficult for the system and their researchers to be unbiased. AI contains societal stereotypes making them vulnerable to clichés. A claim was made to the fact that increased efficiency will lead to the loss of HR jobs. The following assertions were made by some experts that agreed with the positivity of AI applications: text generation could be problematic as large language models are prone to bias; bias that exists in software is perhaps easier to measure and fix than in humans; AI applications used in HR should be smart and care about bias; AI requires rigorous training; it should not replace human decision making; the right use case should be identified and applied to AI.

Q: Do you think there is a risk that AI (algorithms) will introduce bias (exclusion or discrimination) in recruitment/selection decision making?

According to some of the experts, the possibilities of AI is fascinating. However, most agree that AI will reflect existing bias while a few were torn as to whether this was so or not. Data and algorithms can be biased. AI learns from past stereotypical data and reinforces them. For instance, an algorithm trained on existing internet data would prefer male employees over females for a construction job with little consideration for skill. Emphasis was placed on the introduction of diversity bias without human oversight in recruitment and selection. An expert representing an AI association clearly stated that the association was against replacing humans with machines. Another reason cited for the introduction of AI bias was the lack of consideration for diversity, equity, and inclusion.

Biases listed include those related to the training data, the structure of the CV/resume, gender, name, merit/achievements, educational institutions; citizenship/nationality, race/ethnicity/colour/picture of applicant; intangible good qualities where the AI only looks at abilities; address/location say in the north or south of a country; similarity bias where the AI is inclined to hire applicants similar to existing employees. A cited example was that if some companies are 40% more likely to hire a white male than a female, this is reflected in the data and will more likely be recommended by the technology. More reference was made to the algorithm selecting white candidates or preferring men to women such as in the Amazon case. One expert recognized the relationship between gender bias and sexism as well as against the LGBTQ+ community. It should therefore be possible to eliminate factors that create bias without affecting the assessment of people's abilities to hold certain positions.

It was said that there are opportunities to counteract bias if algorithms are based on rules where the system can omit factors likely to influence decision making, here the algorithm would be able to expand the group worked with and increase the possibility for diversity. The issue however is that the algorithm is given access to a large dataset with no set rules making it impossible to decipher how it draws conclusions. It was stated that the output of the model would be wrong if it is not well created by the developer. It is believed that AI models are presently at a stage where they can be analysed with more open-ended questions.

Suggestions were made to receive feedback from candidates to develop AI systems accordingly. Importance was placed on the fair anonymisation of standardized data to mitigate bias. Errors in the system are possible but if the data is correctly fed in, the system is highly likely to produce correct results. An expert pointed out the difficulty in knowing whether the bias introduced by the AI will be more compared to the bias held by regular recruiters. One expert asserted AI mostly discriminates on a racial basis, while humans mostly discriminate based on gender. Some believed the recruiter's role is still essential regardless of their held bias. Another belief was that gender bias could be mitigated or eliminated if more female data analysts and AI developers worked on machine learning. It was acknowledged that AI technologies could be used as tools to correct bias and promote equality if properly designed.

Q: Does your organization use AI (algorithms) in recruitment and selection?

Almost all experts interviewed stated that their organisations do not use AI in recruitment and selection whereas a handful were uncertain of this. Reasons for this being: the scale of the company is too small to require AI; its use is not cost-effective; the desire to process people equally; experience with hiring the wrong people due to AI; lack of AI specialists as opposed to the luck of trust. The few who stated otherwise gave the following reasons as to its use: efficiency; to screen CVs and social media profiles (extracting skills from LinkedIn); to write and post jobs on LinkedIn; to draft emails to candidates; to compare candidates; to collect and filter data on candidates; to create clusters in their database.

Q: What sort of data about individuals do you generally gather for use in the AI algorithms you develop?

Some experts claimed they do not use data on individuals. Others specified they do not collect personal data. One expert explained that their company does not collect direct data on individuals as they work for other companies that do. It was cited that the company created a face-analysis which intercepts the image of the person but formats it to avoid recognition. They work with databases where personal data undergoes a process of anonymization for privacy protection. Non-personalised data and data dependent on other people who provide the data were listed as: those related to skills/training, forensic data; data of people that use certain kinds of services; data on people's habits; data generated by video game system; the number of public transport passengers; shopping desires, frequency and trends; age and address; process based data; past work experience; qualitative and quantitative competencies; certificates held/degrees studied; graduation information; department, gender and military service. In some instance, biological personal data and free available data or open public data were collected but anonymised. One expert claimed that their AI algorithm collects personal data but anonymises it by removing names, addresses, location and background information to only focus on skills to eliminate bias.

A few experts considered gender/sex, degrees studied, statistics on customer behaviour, learning statistics and information about students (specific information not specified) as personal data, which was collected with informed consent, signatures etc. They suggested HR should do the same, as people need to be informed and agree to their data being processed by AI. One expert emphasised the importance of including medical use cases in the discussion of AI bias. It was noted that at the design stage one reaches a better outcome when more personal data is processed. It was also noted that it was necessary to collect data based on intended use and expectation as well as cover all information that will help achieve the best match between applicants and job vacancy. It is however best to keep the AI from showing certain data categories such as gender and age. There was an awareness that this was complicated to uphold in practice given the growing amount of proxy information AI applications process. For instance, one's gender or nationality could be inferred from their name and nationality respectively.

It is important to point out that although some experts claimed they did not collect personal data, data such as addresses are considered personal data under GDPR rules. Similarly, a group of non-personal data that can collectively be used to identify a person are also considered personal data. Also, if de-identified, encrypted or pseudonymised data can be used to re-identify an individual it is still classified as personal data.

Management

Q: To what extent is technology being produced with the purpose of using AI (algorithms) to manage and control individual employees?

Experts reiterated that AI is not widely used in smaller countries/companies but rather in countries with larger companies for instance as a recommendation engine in HR matters. One cited an example that their US branch used AI to monitor employee bathroom breaks and was certain it would not occur in their country. Others thought AI was indeed applicable to this use although they had not encountered such. Some thought to a large extent with AI production for job analytics (performance monitoring and examination) being the most common, while others were uncertain of its use, but thought this may be the case to a small or moderate extent. Some in this category defended its use to monitor employee mouse, mailbox, and online activity rates. Other cited uses of AI for management were the scheduling and managing of personnel work shifts in hospitals during the pandemic; monitoring the behaviour of students during examination; promotion of metrics that need to be achieved although this was difficult depending on the profession. One expert claimed to see an increased uptake of AI in companies to help manage and control employees. This is done to support rather than replace

Funded by the European Union

HR managers and is limited to specific tasks. It was also noted that there are tools for management that did not necessarily include AI. Reason given for the limited use of AI for management and control was the presence of strong trade unions and privacy regulation and that algorithms for this purpose was still under development in some countries.

Given that there is a limit to what the human brain can process there are considerations for AI´s use, as great potential is seen in the training of AI that better understands humans and what affects workplaces in the smartest way. In like manner, one expert agreed that AI could be used efficiently in onboarding or orientation processes, training post orientation, follow up and management as well as scan performance and potential with the purpose to determine career paths. Experts advised that care should be taken when monitoring peoples´ productivity and meting out punishment for low productivity regardless of whether it is done by AI or another technology. The likelihood of the existence of systems to monitor employees in larger companies was established. It was suspected that the laying off of tens of thousands in Silicon Valley was the work of an AI judging who should and shouldn´t be fired.

Q: Do you think that using AI (algorithms) to manage and control individual employees is, in general, a good practice?

Some experts think this is generally a good practice while others outrightly disagree and a few seemed unsure but claimed that using AI for such a purpose raises ethical questions. It's important to use AI for people such as skills improvement as opposed to judging or evaluating them. Those in agreement believe that it improves time management and saves employees time; it can help with specific tasks; it enables managers to eliminate biases associated with ineffectual workers in the team and in turn motivates them. It also helps operational applications to remain stable and is beneficial in automating routine tasks or providing data-driven insights and analytics. Managers can also efficiently use it to create employee development plans. It is a good system to have when measuring productivity and performance in companies where convey or belt work is being managed and in hospitals. It is also useful in industries with limited access such as steel factories to timely warn of dangerous situations or monitor flows in the building and optimise work. Thus, it is dependent on the use case. Care should however be taken in monitoring what people do and with their informed consent; transparency is needed, and personal rights should be respected. Positivity was again based on the condition that there is human oversight and feedback given to employees; AI should not replace interpersonal relationships in management. Some experts thought it was acceptable to take disciplinary action against employees who infringed on company rules and policies although this cannot be directly done by the AI and should be done prior to creating effective communication where employees are aware they are being monitored.

AI does assist managers to manage vast information, but care should be taken here as employees have expressed some concerns as to what data is gathered for the predictive model. What does an employer do when the health information of an employee is flagged? Reasons given for disagreeing with this notion are that surveillance may be good for the company but not employees; instead of managing employees with AI managers should get to know them; AI should be used to improve people and not control them; its usage will result in loss of trust between employees and companies and people should not be put in a box. HR managers were said to be too quick in shifting their burden of responsibility on AI applications; employees receive orders from AI such as with Uber; teams should be made smaller if larger ones are difficult to manage; spending long hours in front of the screen does not necessarily equate productivity as people manage their time differently. One disagreed on the notion that there are other ways of assessing employee productivity.

Q: Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to productivity?

While some experts agree others disagree with the notion that AI should be used to monitor to penalise employees for low productivity. Experts in disagreement thought it was against European law and expressed personal distaste for it. Those in agreement thought it was ok if employees preferred to use it and if there is respect for employee's privacy and dignity. The aim should however be to aid/improve not fire/punish employees. Others expressed acceptance on conditions that AI is only used for its intended purpose, for safety purposes, on reporting-analysis and to measure productivity but with penal action decided by a human. Also, bathroom breaks should not be tracked, and AI should be used to track improved productivity not just low productivity. Other reasons for agreement are that it is useful in adult learning although care should be taken to avoid creating pressure; useful in planning the objectives of performance indicators and creating employee data where information can be accessed for making decisions regarding them.

Others disagreed by claiming that it was ineffective as the knowledge of being observed will lead to employees pretending to have a changed behaviour, thus its use will only restrict worker autonomy without producing reliable outcomes. Discomfort with being monitored was expressed and the issue of trust between employer and employee as well as its use being good for managers not employees was reiterated. It was suggested that the employer should instead create a safe and healthy environment to enable workers' self-determination in freely performing their tasks. Employees being motivated was established as more important.

Q: Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to infractions of companies' rules and policies?

Experts found this practice acceptable, moderately acceptable, and unacceptable. One expert found this acceptable on the premise that disciplinary action is taken by human managers who can better assess the context and nuances of each situation. Some opined AI should be used to simplify people's lives and make accelerated changes needed in the labour market as Gen Z is entering it. Importance was placed on companies not violating their staff. Others asserted that monitoring due to disciplinary motives was dependent on the nature of the company and work carried out as some work with extremely sensitive data, making it acceptable to use AI to ensure compliance. It was also deemed acceptable to use AI to monitor safety/security issues and prevent criminal/illegal acts such as money laundering, embezzlement, fraud, and illegal fishing in the private sector as well as supervise employees slacking on the job, skills management and for small matters. However, AI should not be used to elicit compliance of company values, as such data cannot be used to make automated decisions, also, performance

Funded by the European Union

is not easily measured with keystrokes and screen time.

One expert thought the use of AI for this purpose (to monitor with motives related to infractions of company rules and policies) should be evaluated given the broad nature of company rules and policies. In contrast, another claimed that company rules and policies were clearly defined and thus easier to monitor for infractions with the possibility of prejudice less likely to occur. However, it was still considered an invasive practice and a source of discomfort and stress to employees with suggestions made for it to be used carefully. Again, AI for this purpose was said to require transparency, consensus between parties and awareness on what is fair and unfair to do. It was believed that society had not yet reached this stage. There was also a risk of AI giving wrong results due to a faulty model. It was noted that if a technology (AI) can correct bias and stereotypes, AI solutions could be useful in monitoring productivity, nonetheless this was too soon.

Q: Are there other areas where it is acceptable to use AI to manage and control individual employees?

It was reemphasized that AI tools should be used to aid rather than control people. Importance was placed on the use of AI to create new solutions such as, designing HR applications for detecting what needs to be fixed, changed, or improved; screening for negative behaviours like harassment, plagiarism, unequal pay (based on gender and KPI); controlling public money spending and identifying training opportunities. Social communication among employees as well as profitability and productivity of the company can be enhanced. AI could be used to assist with scheduling and the optimization of workloads and resources and as a self-help tool (one that doesn't share information with the employer) to improve individual productivity. Experts thought it was efficient to use AI to train staff and as a consulting service to present a personal development roadmap, thus, using AI to support employees in determining their career paths or monitoring their productivity along this path. To predict and understand reasons for low productivity and how to fix this.

Q: Do you think there is a risk that AI will introduce bias (exclusion or discrimination) in decision making?

Almost all experts agreed that AI posed a risk of introducing bias in decision making, while a handful disagreed with this notion. One expert could not decide between yes and no because they believed just as the system could develop correct biases, it could also develop false biases. AI decision making is an important issue that requires rules. Experts asserted that the risk was dependent on the training data, correctly inputting the data and the quality of the algorithmic model for the decision-making process that influences outcomes and trends. The risk of bias also comes from monitoring employees using checklist and parameter frameworks. It was explained that although automated decision-making could be problematic, one could use the decision outcome to further investigate, for instance why a person was not at their desk. As pointed out before, the way AI is used today doesn't always make explanations possible. Automation related to AI is hard to pin down. Work was required to prevent and make people aware of this. It was said that one doesn't hear about this issue because it happens unawares as people are not informed of reasons behind decision outcomes. Some examples of bias listed

Funded by the European Union

were those related to gender, citizenship, parenthood, sex, age, social class/background, names, and disability; racial bias e.g., AI used in banks to make decisions on loans; nationality bias e.g., AI used to detect animal photos that may result in bias from photos with different backgrounds in other parts of the world.

Q: Do you think that AI developers take action to screen their algorithms for 'fairness'?

Some experts thought AI developers screen for fairness, others thought they did to an extent but there was room for improvement, while a good number thought they don't do this at all. In agreement an expert claimed that top researchers were meeting to discuss the issue of AI ethics and cited the researcher Joy Buolamwini who initiated the gender shades project after finding out that AI systems greatly misgendered women and darker skinned persons. It was pointed out that although AI developers do this, it sometimes exceeds their expertise or results in a ticking box exercise. It was deemed common for AI developers and HR in large companies to have a fairness checklist. One expert asserted ethics check should be carried out by an independent body like the ALTAI based evaluation.

Experts who disagreed said the algorithm is still screened by developers who harbour unconscious bias; often these teams are also not diverse. There was a lack of awareness on bias as developers only considered bias within the data and not the algorithm or other ethical issues; there is limited understanding on how to screen algorithms for fairness. They are too caught up in the algorithm and other technical things to consider bias, and they regard fairness as a by-product as there is little emphasis and discussion on it. Some argued that AI developers do as they are asked so until this is legally mandated or a standard requirement it will not be addressed. Also, there are no standard debiasing methods. Another claim was the near impossibility of evaluating data for fairness, but improvements are being made for more accurate data although it isn't more just. Constantly following the program and optimizing the system was deemed necessary although a very sensitive issue. While some experts acknowledge the need for raising the awareness of AI developers, others believe they lack interest in rather than awareness on debiasing.

According to experts, history has proven how excited people get about new technology and then tend to forget themselves but eventually developed increased awareness over time. Although the desire to use AI makes one throw in all the data with little thought, there is hope that this will change. Training the model is just one step, it's best if persons involved in the process are aware of the strength and weaknesses as well as biases the data presents. Considered most important is beneficiaries of the results asking about the origin of the data and possible biases. It was believed that there was increased awareness of the necessity to examine other factors besides algorithmic patterns, and many are criticising the implementation of AI applications.

Q: Does your organization use AI (algorithms) to manage and control employees?

There was a consensus that none of the organisations the experts worked at used AI to manage and control employees, except for two where it was used to extract information in one and support management tasks in the other. This was due to the small size of some of the organisations; unreliable and limited spread of the model within the country; the university did

Funded by the European Union

not need nor want the technology as efforts and legislation may overwhelm their HR; its use was considered unethical and inefficient; lack of opportunity in testing the tools; fear of losing competent employees who experience discomfort.

Q: If you were to receive training concerning bias in AI, what topics would you most want addressed?

Below is a list of topics experts would want addressed:

- Ethical concerns and certainty of the need for AI.
- Knowledge on technological history and media portrayal of AI.
- Knowledge on how the technology works and to ensure that people understand these technologies are created by people, thus, the importance of them knowing how bias occurs.
- Eliciting peoples understanding on what they're dealing with before introducing AI into the equation.
- For HR personnel to talk about their own biases, check the statistics and make rules of conduct prior to the large-scale use of AI.
- Understanding the different stages of AI where bias might be introduced.
- Knowledge on how to identify bias in the data source and algorithm as well as good practice/methods for de-biasing and other mitigation strategies.
- Knowledge on the type and quality of data to avoid discrimination and gender stereotypes.
- Accountability
- Awareness of public and industries on biases.
- Knowledge on ethical AI and consideration for ethics particularly in recruitment and management.
- Understanding the consequences of not caring about bias and creating wrong judgements while developing AI.
- Fairness in general and specific to AI.
- Knowledge on data detection and algorithms that can be useful in recruitment processes.
- Knowledge on what constitutes diversity bias in the workplace as well as diverse. available techniques and applications that assist with their mitigation.
- Knowledge on ensuring fairness after deployment of the first AI application.
- Knowledge on existing frameworks, basic concepts of diversity and inclusion.
- Knowledge on existing gender norms and roles as well as software and intellectual property.
- Knowledge on how to create datasets that are fair with regards to recognition and training.
- Knowledge on mathematic instruments and methods to identify the presence of bias.
- Knowledge on testing for validation and automation questions.

Funded by the European Union

6.2.2 HR Executives

The HR executives worked within private corporations/institutions/organisations, a semi-private company, public institutions, private and semi-private research organisations, start-ups including one that develops recruitment and selection platforms, a consulting company, an IT infrastructure company, and a chemical industry operating in the pharmaceutical sector. These organisations/institutions employed from 1 to 9000 workers.

Recruitment

Q: To what extent do you believe that employers or recruitment agencies use AI (algorithms) in recruitment and selection?

Some HR executives reported that AI was used to a small or moderate extent in their organisations or countries while a few could not determine the extent of its use. Big companies were thought to use AI more often than small or medium ones. Its usage was also believed to be more widespread on the international front. Companies in the USA were said to use AI in screening CVs, while a Danish company developed an AI called 'develop diverse' to eliminate words with gender implications. Some of the organisations use algorithms to scan potential candidates on LinkedIn and Star Link. Training the algorithm to correctly search requires a lot of work and time and experts reported not having such time. It was also reported that due to information overload the experiment was abandoned after a year as it was not found to be helpful. The LinkedIn recruiter module works in a similar way but on a different scale.

Instead of AI algorithms, some companies use personality tests, reasoning tests and various supporting data which nevertheless presented technical hurdles given barriers with language and applications not being in a standard format. It was said that people rather than programs were needed to process such applications and have it standardized so the AI can follow a pre-determined criterion, although it was deemed feasible for more simple replacement jobs, particularly low skilled labour. Other experts reported that AI can pre-select candidates when eligibility criteria are entered. End-to-end AI technology was considered unsuitable given today's conditions. Flexibility is required as it is not always possible to find candidates that meet all criteria.

Q: Do you think that the use of AI (algorithms) is in general positive in recruitment and selection?

More experts than not thought AI was in general positive for recruitment and selection as it was evidence based, time saving (e.g., automated CV screening from a large pool of applicants), unbiased from human stereotypes and enforces positive discrimination thereby giving room to the socially marginalised. It also aids in optimization and streamlining. One expert found it hard to say as they believed AI would then have to be able to think like people and not simply read from A-Z. It was further asserted that positivity required preconditions such as eliminating algorithmic bias and using the AI with awareness and prudence. Some experts in agreement were hesitant as they believed it is important what is decided by the AI and what is required of the recruiter. Emphasis was placed on the essential role of transparency, diversity, irreplaceable human oversight and input, and other mitigation measures. Both humans and

Funded by the European Union

AI's can be biased but perhaps it was more controllable with AI. AI should not be used for actual employment decisions; human intervention was required to double-check AI decision outcomes. It is also necessary to evaluate applicant willingness and work motivation during interviews.

Some in disagreement opined that the value of AI was dependent on the training data, developers, and its usage. AI could also cause new biases and discrimination; data set cannot be representative of social diversity. One expert pointed out that because CVs are non-standardised, the AI could overlook something important. Another expert argued that there are more disadvantages than advantages; no one has been able to convince them thus far of the pros. Additionally there are few applications in their domain and thus such tools are not needed. One expert questioned the usefulness of AI systems in the future as since the onset of the covid-19 pandemic applications had decreased. Given that other elements are considered, often, candidates selected do not reflect what is reported in the job description. It was considered useful if the AI could assist with reading non-verbal signals.

In reflection, some experts seemed to have a limited understanding of how AI-based software works. Their answers evidenced that they imagined it often as something that is directly *programmable*, not something that learns based on existing data. There also seemed to be confusion between software in general, and AI-based software.

Q: Do you think there is a risk that AI (algorithms) will introduce bias (exclusion or discrimination) in recruitment/selection decision making?

While almost all experts agreed that AI could introduce bias, one expert refrained from responding due to lack of expertise in AI and the remaining few thoughts AI matched suitable candidates to positions when one considers the definition for behavioural and technical competence expected from the position. Reasons given in agreement were that AI is a black box and humans are not sure how it works; AI developer teams are homogeneous; little is known of the training data and AI assessment; the training data is based on real life in which bias exists: it was dependent on the awareness of the developers on whether the tool they were creating was devoid of human bias. Biases related to ethnicity, nationality, religion, gender, age, language, education, neurodiversity, and disability are introduced into decision making. Experts cited the Amazon case and a scandal involving the Dutch tax Authority. Diversity bias was believed to emanate from language and coding that reproduces these biases. Importance was placed on a blind CV approach; that is, not working with names, pictures, and social security numbers etc., to avoid bias.

AI was acknowledged as being able to pre-select more people than recruiters could. Experts also noted that like AI, recruiters could also influence decision outcomes with their own biases based on gender/sex, language, origins, educational background, and experience such as gaps in CVs when they don't explore the motivation behind this. Users of AI could also configure the system to suit their preferences which could be biased. It was difficult for some experts to fathom the system itself being biased. One expert claimed that the system is programmed to exclude people with non-EU nationalities, and that bias could be introduced into the system in case of lack of maintenance, for instance when a new country joins the EU but is not added to

Funded by the European Union

the list. The potential for bias was also seen when a person was automatically excluded for not possessing a driver's license. Another reflection is that the above examples further illustrate the problem of mixing general software or rule-based approaches and AI software.

Q: Do you know of any strategies to mitigate such bias?

Strategies for eliminating bias were listed as follows:

- Removing personal information such as names, social security numbers, gender, age, date of birth, skin colour, nationality, and pictures from applications (blind CV).
- Conducting research, particularly external research with independent and non-stakeholder researchers.
- Using responsible research and innovation as well as societal readiness as a concept
- Involving the end user and considering gender and diversity factors while developing the technology.
- IT literacy of the people in working with such systems matter a lot.
- When digitalising such companies and onboarding them into AI based hiring solutions, knowing the risks might be useful in thinking of how to mitigate bias.
- Improving and examining the training dataset
- Training the AI systems with better historical data
- Asking human employees to test the AI system during the selection and ensure the diversity of AI solutions.
- Using quotas to ensure diversity.
- Cognitive assessment tests could replace or integrate the deployment of AI applications since, according to one respondent they are more reliable than AI technology (due to the lack of representative datasets)
- Awareness-raising; Training recruiters and AI developers on bias.
- The use of inclusive language
- Human oversight, especially to integrate the outcomes of the AI systems with human empathy. It should be done by people other than cis, heterosexual, white, old, and able men.
- Prioritizing the processing of the soft skills of job applicants rather than their personal information
- Introducing a powerful D&I officer that could, for example, help HR people use inclusive language, broaden their network when disseminating job vacancies, and identify proxy information.
- Transparency, especially to inform people about the deployment of the AI application (i.e., how it works, why it is used, which negative or positive consequences might arise…)
- Development of technologies, such as the de-biaser
- Valorising candidates' behaviours (e.g., exploring how the person concretely put in place the capacities/competences he/she declares to have during interviews).

One expert ascertained that a qualified candidate could be turned down due to geopolitical security considerations.

Q: Does your organization use AI (algorithms) in recruitment and selection?

Majority of the organisations experts belonged to did not use AI in recruiting and selection. Reasons being the AI tool is worse than the standard one; AI's have bias; it is not useful; it is costly; people had more confidence in their judgement; organisations favoured other methods such as blind recruitment and cognitive assessment; not needed due to limited applications received and processes not being digital enough. Some experts explained that recruitment in their organisation was done both through AI and manual interviewing. AI was used as a support tool for candidate screening, case studies, assessments, and competency tests. One company uses an external AI powered solution. One company uses an AI based personality test. One company used it on LinkedIn, but this was discontinued as it was too time consuming. LinkedIn was also used to double-check certain information. Another uses ChatGPT to personize job descriptions, but manual editing is done after this. In a similar vein others use it to check if the job description is written in an inclusive way. One expert expressed that in future their company might introduce automation, not necessarily an AI driven solution. Another was open to experimenting with AI based recruitment and selection tools to experience whether or not it gives good results.

In place of AI, some companies used a digital management software that helps in identifying candidates; a software that detects keywords in the CVs; a system of video-interviews useful for the first phase of screening (candidates send a video interview that will be evaluated by recruiters); a system for the management of data and predictability, in particular, the system would help in understanding if there is an "exit" risk or career advancement aspirations and therefore support in doing targeted actions.

Q: Does your organization use AI for any of the mitigation ideas you mentioned earlier?

According to one expert, successful effort in their organization encompass paying attention to the inclusive language use of particularly hiring managers and holding workshops on unconscious bias and leadership. Another pointed to the diversity and inclusion officer not having extra time to dedicate to this effort as it was an added task.

Q: What sort of data about individuals do you generally gather during the recruitment process?

The following are individual data gathered by organisations:

- Gender, address, education, courses, unions, pension fund, child support, criminal record, sickness certificate.
- Age, nationality, ethnicity (multicultural and 2nd generation immigrants), competence, professional certification, work experience and jobs held, application and CV, online information (to check if candidates have deviations related to the position)
- Date and place of birth, contact details, personal code, telephone number, email, CV format, language skill, available work permit, notice period, how applicants found out about the role, salary expectations.

An expert specified that information about marital status, children, origins, and religion are not asked but sometimes provided by the candidates, while hobbies are asked but are not used in a systematic way. It was explained that expectations were different in blue- and white-collar

Funded by the European Union

positions. In the later information on experience, compliance with job competencies, level of education and knowledge of English language is gathered. In the former consideration is given to the location of the job, due to shift systems, and the presence of chronic illness depending on the physical requirements of the job. Importance was placed on the fit of the applicant to the team.

Management

Q: To what extent do you believe that employers or recruitment agencies use AI (algorithms) to manage and control individual employees?

Almost all experts claimed that AI is not used much in their respective countries. One said it is used to some extent and another said it is used to a large extent in bigger companies while a few don't know to which extent it is used. One country's law on surveillance for the protection of workers was cited as a reason for the limited use of AI. Other reasons were the small size of companies; it was considered unethical; it was believed that HR should directly listen to people without the use of tools. Some uses of AI not related to experts' organisations were for optimization such as managing shifts and holidays for employees who do heavy work like flight crews. Globus AI management tool to organize schedule and staff planning; AI supporting tools to measure commitment and well-being; anonymous tools. These tools were said to be used more often in the private than the public sector.

It was expressed that it is difficult to make AI make human decisions as each case needed to be examined. AI could be used to recognise biased language and act accordingly as well as in the training, performance, and succession planning process, not to manage employees.

Q: Do you think that using AI (algorithms) to manage and control individual employees is, in general, a good practice?

Majority of the experts disagreed that using AI to control workers was a good practice while a few agreed and one said sometimes. Reasons given in agreement were, to help check the staff; the information collected by the AI can be useful in keeping track of employees for salary negotiations; it saves time by reducing the workload of HR managers. It can be used to predict promotions, resignations, and maternity leaves. Experts emphasized the need for people to make the final assessment and for the further development of AI for such use. Others ascertained that it was dependent on the kind of control and management mechanism the AI model was. It was good if it is used for performance management, reduces employee diversity bias, reduces the operational load, and can provide an objective experience when fed with the right data. A question was raised as to whether this would be done. One expert agreed that companies needed to have performance indicators, however, they could find new Key performance indicators (KPIs) to measure efficacy. Another expert mentioned examples from KPI or metrics rather than AI such as, for tracking and alerting based on employees being ill often.

Experts in disagreement expressed employees would focus on 'gaming' the system rather than on actual productivity; the AI has no understanding of external factors that affect productivity such as death in a family or sickness of a child. Biases that exist in the industry remains in the system and perhaps is reinforced by the AI; human oversight is needed; personal interactions like speaking to the manager about burnout are better; it was too intrusive and might pressurize employees; it might take a toll on the wellbeing (especially mental health) of employees; it depicts lack of trust of employees and is unethical. Experts claimed that the success of such tools was highly dependent on HR personnel. The use of the tools is also advertised such that employees will accept them (points to deception or lack of transparency).

Q: Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to productivity?

Some experts believe that it is generally acceptable to use AI to manage and control individuals while others believe that it is unacceptable. AI was deemed acceptable to monitor repetitive work, useful in analysing dataset for connections and measuring employee performance/productivity and deliverables. Acceptance was based on the precondition of people thoroughly understanding AI rather than blindly trusting it. Emphasis was placed on flexible working conditions as such control mechanisms limited productivity. Some thought acceptance was dependent on how AI was used although it would still be difficult and called to question. Nonetheless, AI is helpful with jobs that are measurable on a scale technology can understand, e.g., customer service where one is evaluated on number of calls taken and whether the issues were resolved or not. The expert was unsure whether this was the work of AI or other measuring instruments. It was recommended that care should be taken in order not to negatively reflect on an individual and their experience.

One expert pointed out the importance of extreme safety standards in their organisation as there was a high need to trace and make operations transparent towards the authorities that control them. It is thus useful in assessing sensitive surveillance data to unearth fraudulent actions as well as safeguard and protect personnel in risky workplaces to limit accidents. Those who disagreed said it was extremely intrusive, and synonymous to abuse; it breeds mistrust. It was proposed that such tools could be used as self-help tool for employees to improve their performance without sharing the data with supervisors or HR. It was opined that productivity encompassed other aspects that needed to be assessed by humans, for instance, a condition that prevents one from reading or writing as fast as others, which is not reflected in the productivity number, but an HR advisor is privy to. It was suggested that AI should be used only when there was specific doubt about an employee and not as a one size fits all approach for everyone.

Q: Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to infractions of companies' rules and policies?

Some experts thought this could help in industrial companies or low skilled jobs, but not creative jobs. It was, however, still better to directly approach people. Others thought it was dependent on how and why it is used. It is positive when used to address workers mental health or ensure diversity and used with the discretion of managers but unacceptable when used to excessively exploit productivity. Acceptance was further justified by it being based on laws and regulations, e.g., to enforce that certain websites are not visited with the work computer or that data cannot leave the company devices. Again, when used in regulated areas like production of medical products, as opposed to say a bakery. Some disagreed because it was extremely intrusive and tantamount to abuse. It also shows the lack of trust of employees. It was asserted that workers had a right to not always perform at 100% capacity and thus monitoring could discredit some behaviour that could be useful in improving a process.

Q: Are there other areas where it is acceptable to manage and control individual employees?

Acceptable areas of use that were mentioned were to monitor whether and how parental leave is used; to manage absences; to automatically check for discriminatory language in internal chats or emails; safety, like tracking information on early intervention should something go wrong; big data analysis; jobs where performance was easily measurable like warehouses or factories; on employee pay and work performance, calculating working hours, measuring overall workforce and wellbeing of management. It was considered a step back to use the tech in the service industry as more human oversight is required there. It should be used to support decisions while taking other methods into account to unearth the most suitable way to manage employees.

Some experts expressed fear that their work would become different and may not be as enjoyable as it presently was while others believed these tools to be helpful in making their lives easier with some tasks.

Q: Do you think there is a risk that AI will introduce bias (exclusion or discrimination) in decision making?

Almost all experts thought that there was a risk that AI would introduce bias in decision making. This was because there is a degree of bias in the training data that will result in the repetition and confirmation of existing patterns. Some examples of such biases given include those related to white male dominance, disability, neurodiversity, gender (when a woman is pregnant), sexuality, ethnicity, age, non-native speakers/language issues, origin, health, and religion (Muslims observing Ramadan). It was said that there is a risk if bias mitigation is not properly considered and where logical thinking is necessary; the system might not know or show all the options; the AI system will also eliminate personal aspects. Until people can trust that there would be no errors after outcomes are double-checked the risk remains. Developers were held accountable for biases in the system.

Those who disagreed argued that AI could offer additional support. They opined bias would not be introduced if recruiters understood how AI worked and followed ethical guidelines. A claim was made that since AI tools are programmable, it should be possible to exclude bias (points to a clear misunderstanding of how such tech works).

Q: Do you think that AI developers take action to screen their algorithms for 'fairness'?

Some experts agree and others disagree with this notion, while a handful didn't know. Developers were thought to only think of the outcome; they saw AI as just another tool in their toolbox and a fun experiment; they were unaware of bias as they have not been trained on equality and bias; their teams were not diverse. Again, it was said that fairness was relative to context with no universal definition, thus such screening is not done. One expert thinks it varies

Funded by the European Union

as developers understand AI differently; some want to change the world; others simply think it's cool. Another believes that AI is at the testing phase, but later people will come to think about its limitations. Also, one needs to be more critical about justifiable and beneficial use. Some experts in agreement were unsure and rather hoped that this was done. They believe this should be discussed between developers and users, and that HR managers had a responsibility to verify these tools. It was suggested that an ethics board should be present during the development of such software. One expert asserted that AI developers should take steps to screen their algorithms for fairness.

It was expressed that it was ok if AI is developed based on concepts of justice acceptable within the society or organisation. The use of AI in many fields make it difficult to imagine a world free from discrimination, but the type of data and the field in which it is used is important. For instance, ethics exists in company management but not in social media. With time the development of systems will progress. It was noted that the concept of fairness was vague, and our biased society requires a cultural change. Both AI and people are biased; people make decisions based on algorithmic data.

Q: Does your organization use AI (algorithms) to manage and control employees?

Almost all the experts' organisations do not use AI to manage and control employees. Lack of use pertained to the small size of companies and level of development of HR departments, including non-digitalised processes; all the possibilities of AI were yet to be revealed; fear of their reputation being tainted; not cost effective; not presently needed and no specific reason.

One organization uses "pulse measurement" (a dialog-based tool to ensure employees' well-being at work). If an employee(s) gets above a certain score on the survey, the algorithm will ask a new question until the issue is detected. If there is a red flag for mental health the manager will be told that measures must be taken. One production company used it to measure the instant performance of employees. It is used to scan academic studies over the internet and scans citations as well as views under employees' profiles. The google citation scanning system was cited as a similar system. One organisation uses an App for time management.

Q: If you were to receive training concerning bias in the use of AI in human resource management, what topics would you most want addressed?

Topics experts mostly want addressed are listed below: (two experts were uncertain)

- To develop understanding of how the AI system functions, its uses, pros, and cons as well as to know who develops the AI.
- Comparing findings from technological research education to see where society stands with regards to AI presently.
- To ensure neutrality, through understanding of AI systems implementation and create an analysis of the diversity and inclusion sought after in candidates.
- Knowledge on diversity and other biases as well as those reflected in AI technology. Employing for instance role play to raise awareness and giving concrete examples and experiences about prejudices related to AI.

- How to address these biases
- Knowledge on optimization results
- Knowledge on the ethics of the selections
- How to process personal data more efficiently and adequately, to better match the job applicant with the job vacancy.
- How to use inclusive language
- How to use LinkedIn and other social media in a responsible way
- How to define diversity biases in the workplace, especially the unconscious ones and the socio-ethical impact they might have.
- Knowledge on AI frameworks applied to HR, which are the rationales and data that are at the basis of AI development, how to read and interpret data that emerges from the analysis made by AI.
- Knowledge on available AI tools and their functionalities: what stages it will involve, how it works, how they mitigate bias, a demonstration, drawbacks, and price.
- How racial bias can be mitigated.
- Ethical use of AI and legal requirements
- How AI works, and how to train it.

It was recommended to have a standard for AI that was more just and fair; individuals should move away from prejudiced thinking. AI was considered by some experts as a system that will eliminate prejudice.

6.2.3 Summary

This report highlights the perception of AI and HR experts on bias pertaining to AI solutions and their use in the recruitment and management of employees. There was a consensus on the reflection of societal bias in the training data set which was said to result in the AI algorithm repeating and reinforcing these biases. AI was reported to have inherent bias emanating from a faulty model, the margin of error and input parameters. There was a risk of both AI developers and HR recruiters having unconscious bias that could affect the outcome of algorithmic decisions. Diversity bias related to gender, race/ethnicity/nationality/origin/skin colour, age, and disability which were easily noticeable was reported as the most common form of bias in recruitment. Majority of the organisations´ experts were employed at did not use AI to control or manage workers for any reason. This was due in part to protection and privacy regulations, strong trade unions, desire for fairness, and cost. AI was also thought to be commonly used in bigger companies and larger countries particularly the United States than smaller companies and countries. Most experts also disagreed with the use of AI for such purposes. They opined that AI should be used to assist rather than control workers, and emphasized the importance of human oversight as there were crucial parts of managing people that was beyond the scope of AI. However, most believed that AI was positive when used to assist employees and employers to varying degrees. The use of AI was accepted on conditions of safety and security, transparency, human oversight, and explicability in recruitment processes, as well as to improve employee productivity. Some mitigation measure suggested for reducing bias during development of AI models and their use in HRM include developing a bias mitigation system such as the de-biaser; having diversity quotas during recruitment;

Funded by the European Union

raising awareness on bias; examining and anonymising the training data; transparency and human involvement. Experts commonly expressed that they preferred training on AI ethics, diversity and inclusion, bias, and mitigation strategies for bias.

Funded by the European Union

7 The survey

The BIAS survey currently seeks to map the personal attitude of job applicants and workers towards diversity biases of AI applications in the labor market. In particular, the BIAS Consortium expects to gain the personal experience of job applicants and workers who have interacted with AI applications in the labor market to understand how this interaction has been perceived to work and has made them feel, besides their general opinion on the matter. For the survey to be as inclusive and representative as possible, the BIAS Consortium aims to reach 4000 respondents across the European Union, in addition to Iceland, Norway, Switzerland, and Türkiye. Also, it finishes with some demographical questions that allow the BIAS Consortium to determine the degree of diversity that characterizes the pool of respondents. All this information is fully anonymized. Although it is still possible to fill the survey out, the *interim* results prove that respondents generally have a neutral attitude toward AI applications in the labor market. Age and nationality are referred to as a common ground of discrimination, and a lack of transparency about the use of AI applications exists. Fairness of AI applications is perceived to overlap with non-discrimination and procedural guarantees.

7.1 The drafting process

Based on the first research outputs of the literature review, ULEID drafted the first version of the survey in February 2023. In doing so, it also considered the semi-structured template for the expert interviews to create additional synergy between T2.2 and T2.3 and inform the design of the Debiaser with more comprehensive research outputs. ULEID decided to divide the structure of the survey into the following four sections:

1. Interaction, where the BIAS Consortium seeks to understand whether the respondent has ever interacted with an AI application in the labor market and, if so, how it has worked.
2. Experience, where the BIAS Consortium asks the respondent to share how the interaction with the AI application has made them feel.
3. Perception, where the BIAS Consortium maps the more general attitude of the respondents towards the deployment of AI applications in the labor market.
4. Demography, where the BIAS Consortium collects some personal information about the respondent to ensure that the pool of respondents is as diverse and inclusive as possible.

This structure is currently explained in the introduction of the survey, which also clarifies its reasons and aims, its anonymous nature, its length, and the chance for the respondent to get in contact with the BIAS Consortium via e-mail to provide feedback and ask questions.

The first draft of the survey was internally reviewed by all the Consortium partners, except for LOBA and Crowdhelix, who are not part of this project activity. Whilst there was a common agreement on the structure, aim, and content of the survey, the Consortium partners discussed how to ensure that the language would be:

- **Gender-neutral and inclusive**, *e.g.*, through the use of the third person plural.

- **Clear, simple, and accessible**, *e.g.*, the term 'bias' was replaced with the periphrasis "prejudices and other unsupported judgements […] because of their personal characteristics". Some definitions, such as in the case of negative and positive discrimination were given.
- **Adequate, respectful, and culturally sensitive**. In particular, the Consortium partners extensively discuss the use and agreed on the exclusion of the word 'race' that, in some European countries, is not considered socially acceptable. Also, it was reported that the collection of personal information about sexual orientation could be complicated in European countries where homosexuality is still criminalized or highly socially stigmatized. In this case, the question has been maintained since the respondent always has the chance not to respond. Lastly, every demographic question is formulated in a way that allows the respondent to self-identify.
- **Equivalent.** For instance, national educational systems vary, thereby leading to different classification systems and language nuances.

The BIAS partners consulted with their internal network of diversity experts to agree on these issues. They checked how the European Commission and other EU institutions and projects have formulated their own surveys. Furthermore, it was decided that the survey's target audience would be any job applicant or worker over 18 years old who has interacted with an AI application in the labor market.

The final version of the survey was approved in March 2023, when the translation process could begin. Indeed, for the dissemination of the survey to be as widespread, inclusive, and effective as possible, the BIAS Consortium decided to translate the document into all the EU languages in addition to the official ones of its non-EU partners *(i.e.*, Icelandic, Norwegian, and Turkish). All the translations were done by the Consortium partners and their professional or personal networks. When possible, the translations were double-checked by another native speaker or a proficient language user. A disclaimer was included in the introduction to the survey to let the respondent be aware that the translation was not professional while apologizing for any possible mistake or inaccuracy. The chance to report any possible mistake or inaccuracy was given by contacting the BIAS Consortium via e-mail. While the translation process was ongoing, NTNU sent the final version of the survey to the Norwegian Data Protection Authority to get the green light at the beginning of April 2023. More specifically, the Norwegian Data Protection Authority acknowledged that the survey had been created to guarantee its respondents' anonymity. For this purpose, the public body also checked the data protection agreement between ULEID and Qualtrics, namely the survey tool the BIAS Consortium decided to use because of its privacy and user-friendly interface.

After receiving the data protection clearance, ULEID started to upload the survey and all the translations on Qualtrics. This procedure took more time than expected since Qualtrics repeatedly failed to save the translations and/or continued to overwrite the text. Simultaneously, ULEID had bilateral meetings with LOBA and CROWDHELIX to discuss the dissemination strategy described in the following subsection. The survey was officially launched on the first week of May, and its final version is attached to this deliverable as Annex VII.

7.2 The dissemination strategy

Funded by the European Union

In collaboration with ULEID, LOBA developed and adopted the communication and dissemination strategy of the survey that briefly involves the following actions:

- A dedicated page on the BIAS website.
- Several templates for images, with the content in the native language of each Consortium partner.
- A paid campaign on Facebook, Twitter, and LinkedIn, where the BIAS project has its own social media accounts.
- The publication of a post on the Crowdhelix platform.
- The registration on SurveyCircle, namely a software designed for recruiting participants through mutual support. In a nutshell, the more the BIAS Consortium fills other people's surveys out, the more other people will take part in its studies.[12]

Additionally, since the launch of the survey, each Consortium partner has shared the link to the survey via e-mail. In addition to specific stakeholders that might have been interested in filling it out, the Consortium partners got in contact with other consortia (*e.g.*, FINDHR, AEQUITAS), as well as relevant networks (*e.g.*, the Dutch AI coalition and the European Trade Union Committee for Education). ULEID and other Consortium partners with teaching commitments asked their students to help with the dissemination of the survey and the filling out. Most Consortium partners included the QR code of the survey in their PowerPoint presentations, when joining national and international conferences (*e.g.*, the Gervigreind, siðferði og samfélag (Artificial intelligence, ethics, and society) conference that took place at the University of Iceland on 5 June 2023, as well as the European Workshop on Algorithmic Fairness that took place in Winterthur on 7-9 June 2023). All the participants to the co-creation workshops will have some time to fill out the survey.

7.3 *Interim* research outputs

At the time of internally revising (*i.e.*, the end of June 2023), the survey got 665 respondents. Although the BIAS Consortium plans to reach the KPI of 4000 respondents by early autumn and will include a full and detailed data analysis in the second iteration of this deliverable that is due in M15 (*i.e.*, January 2024), it is already possible to present some preliminary results. In this way, this section can somewhat contribute to the conclusions of this deliverable, as well as can identify some mitigation plans to ensure that the pool of respondents is as diverse and representative as possible.

7.3.1 Demographic information

The pool of respondents is already geographically diverse, in the sense that the survey has already reached people who are located in all the EU Member States, as well as Iceland, Norway, Switzerland, and Türkiye. Simultaneously, all the same European nationalities but Finnish, Latvian, Luxembourg, and Maltese are covered. Because the multiple choice also included the

[12] More detailed information about the dissemination strategy is included in D7.2, namely the report on dissemination activities.

Funded by the European Union

option 'Other', it is worth observing that 66 respondents are located in Europe but are third-country nationals. Also, 22 respondents refrain from sharing this demographic information.

More than half of the respondents belong to the group age between 29-39 (51,56%), with the ones between 18-28 and 40-50 being equally represented and respectively amounting to 20,11% and 19,55%. Unsurprisingly, respondents between 51-61 and over 61 are less represented and correspond to 4,25% and 2,67%. The reason behind this disparity is likely to lie in a generation gap, considering the recent and growing design of AI applications in the labor market. The response rate of people who prefer not to answer is 2,27%.

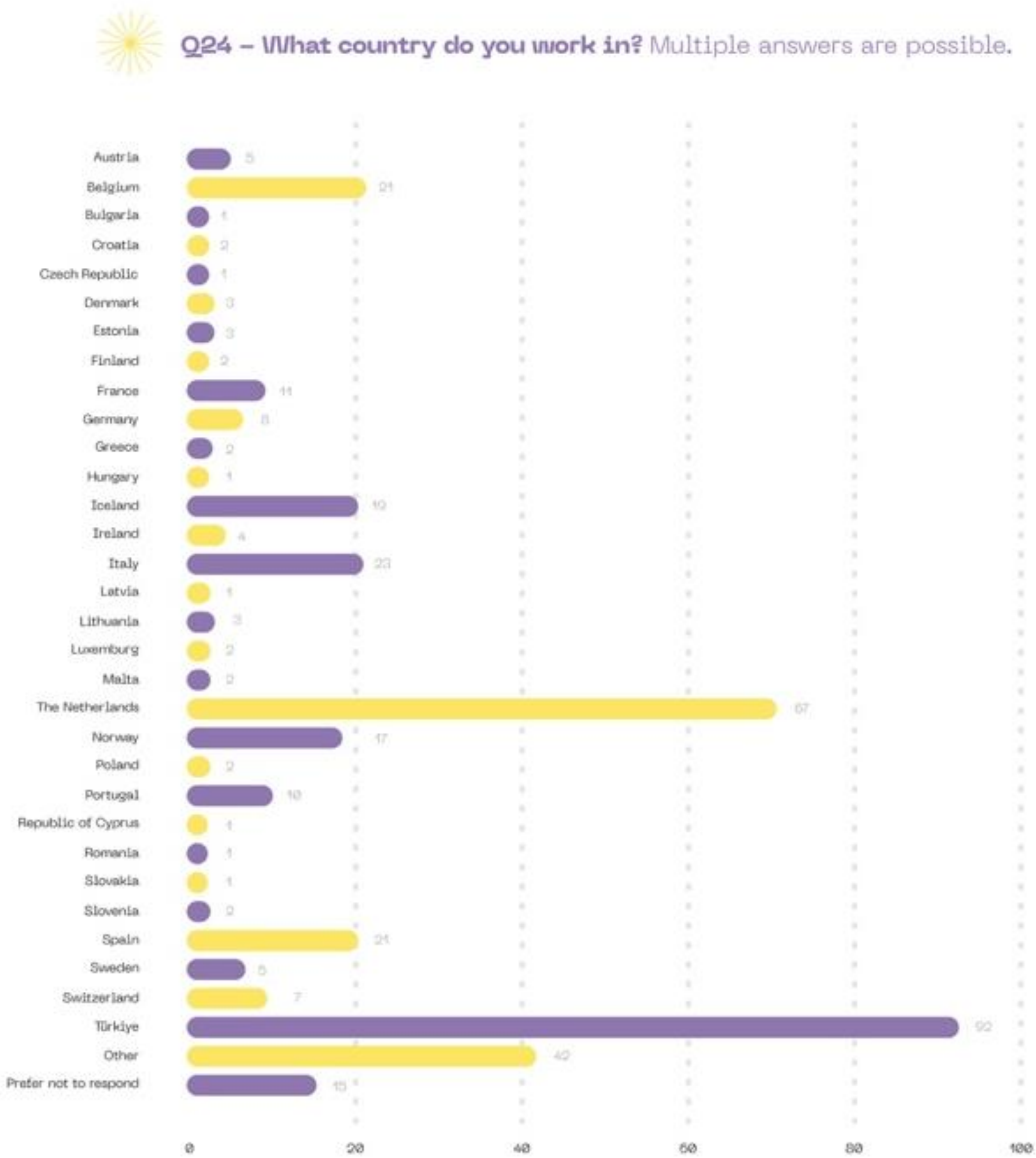

Figure 6: Where survey respondents work

In terms of the level of completed education, the majority of respondents have a bachelor's (29,18%), master's (38,53%), or PhD (20.96%) degree. Respondents with primary school, secondary school, and vocational training education respectively amount to 1,13%, 2,83%, and

3,40%. The response rate of people who prefer not to answer or believe that the list of options does not reflect their self-identification is always 1,98%.

Whilst 5,06% of the respondents prefer not to indicate their gender, women and men are nearly equally represented and respectively compose 44,94% and 47,75% of the overall respondents.

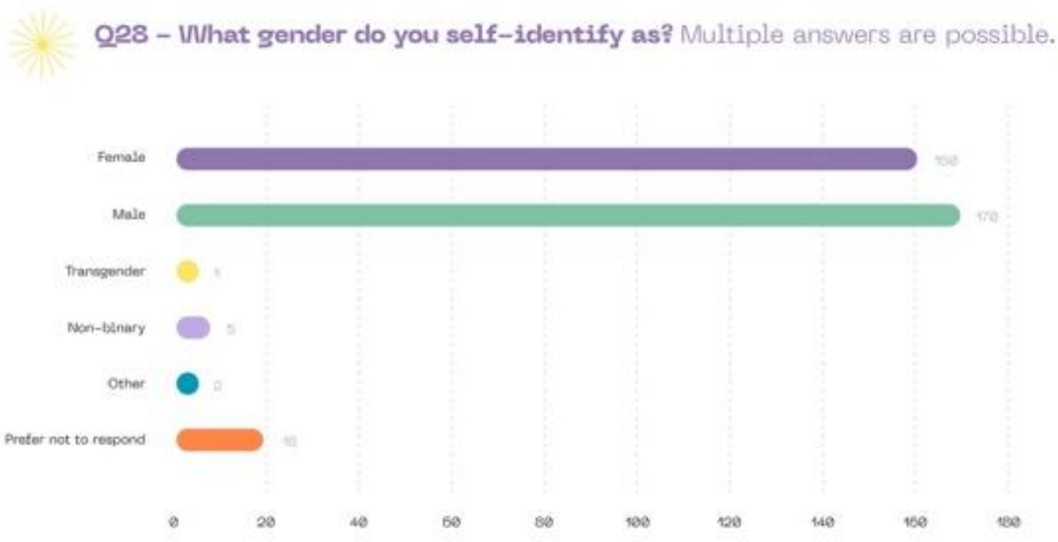

Figure 7: The gender of the survey respondents

Besides, some transgender (0,28%) and non-binary (1,40%) people filled the survey out. 0,56% of the respondents indicated that this list of options does not match their self-identification. The response rate of people who prefer not to answer is 5,02%. To ensure more diversity, the BIAS Consortium is currently trying to get in contact with national and European LGBTIQ* groups.

Although 17,28% of respondents do not indicate their sexual orientation and 2,55% of them self-identify with a sexual orientation other than the ones listed in the survey, straight people are the biggest portion. Straight people, indeed, amount to 66,01% of the overall respondents, with lesbian, gay, bi-sexual, and a-sexual people respectively reaching 1,42%, 5,95%, 6,23%, and 0,57%. As previously said, the BIAS Consortium plans to directly involve LGBTIQ* groups, to ensure more balance in gender and sexual orientation.

When asked about the self-identification of race and ethnicity, 55,01% of the respondents consider themselves white. Also, a high amount of people prefers not to answer (13,28%) and believe that the close list of the survey did not represent them (5,42%). Moving to the other, specific race and/or ethnicity groups, 8,67% of the respondents self-identify as Asian, 7,32% of the respondents self-identify as Latin American, 4,61% of the respondents self-identify as Middle Eastern, 2.17% of the respondents self-identify as black, 1,36% of the respondents self-identify as North African, 0,54% of the respondents self-identify as Roma. Also, 1,63% of the respondents considered themselves to belong to a multiple racial/ethnic group. Accordingly, 60,06% of the overall respondents do not believe to be part of a minority group in their country, with 31,16% of people replying in the affirmative and 8,78% preferring not to respond. To reach a diverse pool of respondents more effectively, the BIAS Consortium is engaging in extensive consultation with civil-society organizations and research centers working on the topic. For example, a representative of the Racism and Technology Center will join the Dutch co-creation workshop on 4th July 2023.

Similarly, the BIAS Consortium is trying to get in contact and collaborate with civil-based society organizations, networks, and other entities working with people with physical and/or mental disability since 89,80% and 90.37% of respondents respectively self-identify as able people. For example, a ULEID has recently been in contact with the European Disability Forum, which will also participate in its co-creation workshop on 4th July 2023. When asked about their physical and mental disability, 3,40% and 3,83% of the respondents respectively answer in the affirmative, whilst 6,80% refrain from answering.

Although trade union representatives have been involved in the BIAS project since the beginning, it is interesting to observe that 78,19% of the respondents are not affiliated with a trade union. The response rates of people who reply in the affirmative or prefer not to reply respectively are 16,43% and 5,38%.

The percentage of people describing their political opinion is extremely diverse. People with a strong political opinion and those who are not affiliated with a political party nor have political opinions are equally represented and respectively amount to 24,65% and 24,93% of the overall respondents. This result is followed by respondents who preferred not to answer (18,41%), publicly share their political opinions (12,46%), are affiliated to a political party without a formal membership (5,95%), and are affiliated to a political party with a formal membership (3,97%).

Regarding religious self-identification, 24,08% of the respondents consider themselves atheists and 15,30% consider themselves agnostics. The response rate of people who prefer not to reply is 16,71%. 16,71% of the respondents self-identified as Muslim-Sunni and 1,42% self-identified as Muslim–Shia. Residually, 11,33% of the respondents self-identify as Catholic, 3,40% of the respondents self-identify as Protestant, 1,98% of the respondents self-identify as Orthodox Christian, 0,28% of the respondents self-identify as Jewish, 0,57% of the respondents self-identify as Buddhist and 1,42% of respondents self-identify as Hindu. Interestingly, 6,80% of the respondents believe that the close list of possible answers was not representative of their religion.

7.3.2 Interaction with AI applications

Turning to the respondents' personal experience with AI applications in the labor market, it is necessary to clarify that the survey begins with three questions aimed at making the doubtful respondents reflect on their past experiences. Out of the 22,94% of the respondents that were not sure to have interacted with an AI application in the labor market, nearly half changed their reply and agreed to continue filling out the survey. Otherwise, they could directly move to collect the general attitude towards AI applications in the labor market.

In the attempt to understand where AI applications are mostly used, most respondents (52,69%) have worked in the private sector. While the remaining 37,63% belongs to the public sector, 9,68% of the respondents surprisingly reply ‘Other’. When asking respondents to clarify the

Funded by the European Union

sector in which they work potentially, 18.83% of the respondents believe that the options listed in the survey do not fully match their experience. Most respondents respectively work in education (22,40%), IT (12,98%), and manufacturing (10,84%). These results are followed by healthcare (4,71%), financial services (4,56%), culture (3,14%), transport and storage (2,14%), insurance services (2%), construction (1,28%), social welfare (2%), food (1,43%), wholesale and retail (1,28%), accommodation (1,14%), agriculture, fisheries, and forestry (1,14%), sport (1,14%), real estate (1%), energy supply (0,86%), waste management (0,71%), water utilities (0,57%), mineral extraction (0,57%). Additionally, 63,90% of the respondents are employed permanently. Instead, 19,92% of the respondents are employed temporarily, 6,85% are self-employed, and 3,11% are still trainees. The remaining 6,22% think the possible answers do not reflect their experience.

When describing the AI applications they have interacted with, respondents mostly believe that the technology analyzed their personal information (19,94%) or put them in contact with possible employers (18,48%). Alternatively, the AI applications are considered to test the skills of the respondents (8,77%), inform them about some job practicalities (8,14%), train their skills (6,58%), check their time availability (6,78%), rank them (6,78%), monitor their productivity

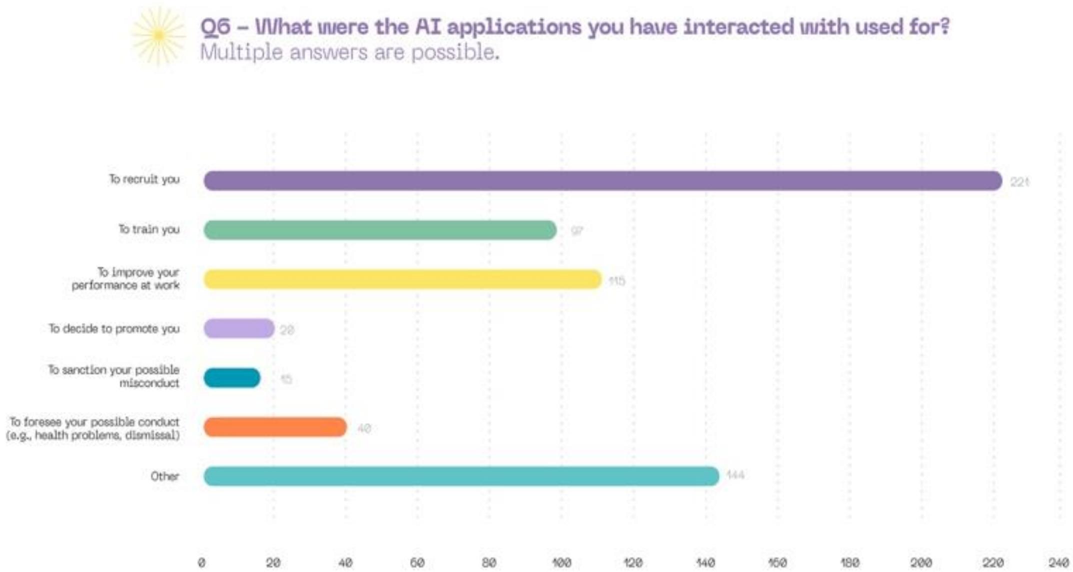

Figure 8: The purpose of the AI applications used by the survey respondents.

(5,22%), replace a human being in a job interview (3,55%), and check the health conditions of the respondents (3,65%). According to 12,11% of the respondents, they have interacted with an AI application that performs tasks other than the one previously mentioned. Regarding the purpose of the AI application, 33,90% of the respondents report that it has been used to recruit them. This purpose is followed by improving the performance at work (17,64%), training the worker (14,88%), foreseeing possible problems related to, *e.g.*, health and dismissal (6,13%), deciding to promote the worker (3,07%), and sanctioning possible misconduct (2,30%). However, 22,09% of respondents believe this list of purposes does not match their personal experience. Besides, most AI applications (38,34%) are seemingly designed to interact with the

Funded by the European Union

respondents. Others (20,38%) aim at deciding about the respondents or helping another person (20,03%). 21,24% of the respondents report reasons other than the ones mentioned.

To understand whether and to what extent AI applications in the labor market are transparent, the survey includes some questions about the information received about the use of the AI application. In this regard, nearly half of the respondents (46,68%) declare not to have been informed. Instead, 30,91% of the respondents do not remember, and the remaining 22,41% of the respondents answer in the affirmative. When informed, it appears that this communication has mostly occurred before the interaction with the technology (60,38%) and in the written form (56,44%). Instead, 23,58% of the respondents say that information has been given before and after the interaction, whilst 10,38% of them report that it only happened before. The remaining 5,66% of the respondents do not remember. In terms of communication means, the written form is sometimes accompanied with the oral one (28,71%). Only 9,90% of the respondents say to have only received oral information, whilst the remaining 4,95% do not remember. As far as the content of this information is concerned, most respondents report that they have been informed about data processing (29,49%). Alternatively, the information covers the way the AI application works (24,88%), the reason behind the use of the AI application (21,66%), the chance to get further clarification on the use of the AI application (11,06%), and the possibility object to the use (9,22%). The remaining 3,69% of the respondents does not remember. Of significance is also that 59,06% of the respondents do not know whether there was a human being involved in the deployment of the AI application. In comparison, 23,49% of the respondents believe that it was fully automated, and 10,74% of the respondents believe that a human being could decide whether, when, and how to use the AI application on a case-by-case basis, and 6,71% of the respondents believe that a human being monitored each decision made by the AI application.

When asked whether they could object to the use of AI applications, most respondents (36,22%) answer in the negative. This reply is followed by people who could not remember (23,30%), could object (20,71%), and could sometimes object (19,74%).

7.3.3 Personal experiences of fairness and diversity bias

When asked whether they felt comfortable or not with the use of AI applications in the labor market, most respondents report that their experience has been neutral (36,74%). In addition to people who prefer not to respond (6,08%) and others who think that the list of options does not reflect their experience (3,41%), the said, neutral reply is followed by respondents who felt uncomfortable (18,49%), slightly comfortable (17,27%), very comfortable (11,44%), and very uncomfortable (6,57%).

The survey continues with a question about the AI-driven collection of personal data. The decreasing percentages are as follows: education (11,44%), gender (11,09%), age (10,86%), previous work experience (10,74%), languages spoken (9,29%), nationality (7,90%), address (6,33%), availability (5,92%), race and/or ethnicity (3,95%), family status (3,31%) disability (2,67%), physical health (1,86%), sexual orientation (2,32%), mental health (1,45%), gap years (1,74%), religion (1,28%), trade union affiliation (0,75%), genetics (0,64%), and political affiliation (0,64%). According to 5,81% of the respondents, other personal data have been collected.

Funded by the European Union

In investigating the personal experience of negative and positive discrimination, nearly half of the respondents feel that they have been negatively discriminated against due to the collection of previously mentioned personal data (51,58%). Instead, 18,98% believe this discrimination has not occurred, and 29,44% cannot say. When asked whether negative discrimination has occurred due to the collection of personal data other than the ones previously mentioned, most respondents were not sure (45,50%). In comparison, 32,12% of the respondents answered the negative and 22,38% answered the positive. Some respondents further explain their reply and argue, for example, that their hobbies have been asked and quotas could have prevented them from being recruited. Overall, great emphasis is put on nationality and age. Gender, sexual orientation, and language are also mentioned. Regarding the positive discrimination that AI applications could promote, 45,75% of the respondents are not sure, with the remaining 34,75% and 19,50% of them respectively answering in the affirmative and negative. A similar margin of doubt also arises from the possible collection of information other than the personal data previously mentioned (49,25%), with the remaining 28,75% and 22% of the respondents answering in the negative and positive. Again, hobbies are considered a piece of information that could lead to positive discrimination. Interestingly, one respondent feels that domestic and care work can favor certain job applicants. Some respondents admit that their privilege (*i.e.*, being white, young, and well-educated) could favor them after assuming the AI application has been trained on a similar dataset.

Against this backdrop, the survey asks respondents whether collecting all this personal data and proxy information could be important to perform the job tasks adequately. In this regard, 35,35% of the respondents answered negatively, and 34,09% cannot say. The remaining 30,56% of the respondents replied in the affirmative.

In terms of the personal experience of negative discrimination arising from disclosing all the said personal data and information, 40,40% of them cannot say. Instead, 38,89% of the respondents answer in the negative, and 20,71% in the positive. When further explaining their replies, it is worth observing that one respondent indicates that their name, which has a Christian origin, has often led to their negative discrimination in a country where most of the population is Muslim. Similarly, another says to have changed their name to be more easily socially accepted by Belgian employers. Generally, great emphasis is put on age and nationality. Race is sometimes mentioned. A similar pattern of answers arises from the corresponding question about information disclosure and positive discrimination. In further clarification, some respondents regard their female or male gender as an indicator. On the one hand, it appears that women increasingly benefit from gender quotas. On the other hand, men still consider themselves to be the most privileged social group, thereby being favored due to, e.g., less domestic and care work. Past working experience, age, and hobbies are also mentioned.

7.3.4 General attitudes towards fairness and AI applications in the labor market

Moving to the general attitude towards fairness and AI applications in the labor market, the majority of the respondents (30,96%) do not have neither a positive, nor a negative attitude. Similarly, 24,38% and 25,21% of the respondents have a slightly positive attitude or a slightly negative attitude. The remaining 6,58% and 4,38% of the respondents respectively have a very positive and a very negative attitude, whereas 5,21% of the respondents cannot say and 3,29%

Funded by the European Union

believe that their personal opinion is not reflected in the close list of options. Respondents are then asked to rank from 1 (*i.e.*, very positive) to 7 (*i.e.*, very negative) the impact that the deployment of AI applications might have on the labor market. Based on the feedbacks received, it appears that some respondents have not understood or have agreed with the formulation and/or design of this questions. In any case, the results are as follows:

1. Productivity.
2. Social inclusion.
3. The transparency of the decision-making.
4. Data protection.
5. The accountability of the employer.
6. The mental well-being of the worker.
7. The personal autonomy of the worker.

Additionally, respondents are asked whether they believe other positive or negative impact could arise. Amongst the answers, it is worth observing the chance for the employer to identify technical and diversity bias when monitoring the functioning of the AI application (positive), the cost effectiveness (positive), and potential job loss (negative). One respondent mentions fairness without further defining the concept.

The following question, though, explicitly asks respondents to choose their preferred definitions of fairness, the results being as follows:

1. When it does not discriminate against one or more people because of their personal characteristics (e.g., gender, age, disability): 19,43%
2. When it is possible to explain how the AI application has made a certain decision: 14,87%
3. When it is possible to challenge the decision made by the AI application: 13,38%
4. When the worker knows how the AI application works: 11,29%
5. When a human being monitors how the AI application works and reviews its decisions (11,43%)
6. When the AI application allows the individual to work in a way that suits them (8,07%)
7. When it can perform better than a human being working in HR (7,92%)
8. When it treats similar people (*e.g.,* all Black women) in a similar way (5,23%)
9. There will never be a fair AI application (3,44%)
10. When it favors one or more people because of their personal characteristics (3,21%)
11. Other: 1,72%.

In this context, it is interesting to observe that great emphasis is put on an understanding of fairness that correspondents to the definition of non-discrimination from a negative angle provided in Section 2.1, which is then followed by procedural elements of fairness including transparency and explicability. Instead, positive discrimination is not commonly perceived as an indicator of fairness.

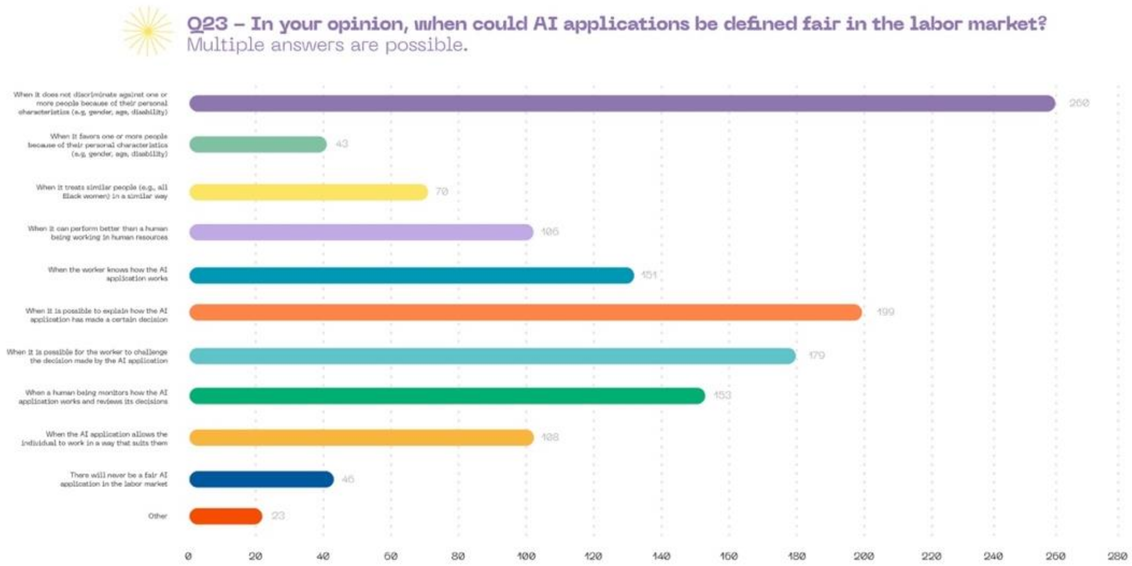

Figure 9: The definition of fairness of the survey respondents

Funded by the European Union

email
info@biasproject.eu
website
www.biasproject.eu

8 Discussion and conclusion

This deliverable has explained how the BIAS Consortium has engaged in desk and empirical research, the aim being to gain consolidated knowledge about AI applications and fairness in the labor market and lay down the basis for the design of the Debiaser. Accordingly, Section 3 began with a scoping review that provided a state of evidence in the field of AI-driven recruitment and selection procedures and focused on competing formulations of fairness and bias, along with legal and extra-legal measures to ensure and address them, respectively. Section 4 then moved to the creation of the national labs, namely the creation of a pool of diverse stakeholders (*e.g.*, employees, employers, HR practitioners, AI developers, policymakers, trade union representatives, and scholars) that could contribute or be interested in the implementation of the BIAS project, because of their professional expertise and/or experience. Section 5 explained the mapping exercise, which attempted to scope the current design and/or deployment of AI applications for HR management. Section 6 summarized the expert interviews, where the BIAS Consortium focused on the professional experiences and personal attitudes of HR executives and AI developers towards diversity biases of AI applications in the labor market. Section 7 provided a preliminary analysis of the research outputs of the survey designed to reveal the fairness perceptions of job applicants and workers when interacting with AI applications in the labor market. Considering the common goal and the resulting synergy between all these project tasks, what follows is a brief assessment of the research findings given the said participatory development of the Debiaser.

The literature review, the mapping exercise, and the expert interviews demonstrated that, although we do not yet know the precise extent of the use of AI applications in the labor market, it is possible to divide them into two main categories. On the one hand, some technology primarily aims at facilitating and/or replacing HR practitioners in the implementation of specific tasks, such as selection and recruitment, time management, and performance assessment. For example, an AI-driven tool could pre-screen job applications and find the best match between job vacancies and job seekers. On the other hand, other AI applications attempt to identify and mitigate diversity biases that negatively affect HR practices, be they manually or automatically done. This could be the case of any technology reviewing gender-neutral and inclusive language when drafting a job vacancy or interacting with job seekers.

Also, the same research activities proved that the design and deployment of AI applications in the labor market carry ambitions and limitations, especially in the recruitment and selection process, which corresponds to the material scope of the BIAS project. At first sight, AI applications could identify, attract, screen, assess, interview, and coordinate with job applicants more effectively because of their ability to process information and make decisions at volumes and speeds that far exceed human capacity. However, they can simultaneously reproduce and perpetuate diversity bias, leading to discriminatory and harmful consequences for job seekers and society. Said otherwise, it appears that most AI-driven recruitment and selection procedures are far from being regarded as fair.

Fairness is a frequently used term yet a concept difficult to define precisely. In this regard, the literature review allowed the BIAS Consortium to distinguish between the main definitions and categories of fairness in philosophy, social science, and law to be read in conjunction with the technical working paper currently produced by NTNU and BFH. No matter how it is framed, however, it appears that the very idea of fairness aims at balancing a power asymmetry. This also emerges from research targeting the perceptions of fairness of job applicants and their main reliance on procedural fairness, namely compliance with a set of legal and ethical requirements that enable each job applicant to benefit from the same opportunities. *Inter alia,* these requirements could cover transparency, explicability, and human oversight. Such a research output is similarly confirmed by the survey's preliminary results, where great emphasis is likewise put on non-discrimination. At the same time, the literature review and the expert interviews show that, for the HR community, fair recruitment and selection overlap with the best match between the job vacancy and the job applicant to the most significant benefit of the organization. Although diversity is considered necessary, its transposition into practice is seemingly complex. For this reason, positive discrimination is seldom supported, as also emerging from the survey's preliminary results.

Perhaps, the perceived need for procedural fairness lies in the contemporary absence of transparency of AI applications in the labor market. While the mapping exercise demonstrated that it is complicated for people to understand how technology is designed and used, the survey's preliminary results explicitly report that job applicants and workers have rarely received information about the purpose and functioning of the AI application. For this reason, they could also not say whether and to what extent another human being was involved in the AI-driven HR practice.

In construing fairness concerning its opposite, the literature review and the expert interviews focused on the definition and the creation, identification, and mitigation of bias. With the formulation of diversity bias varying in the literature, there is some common agreement on their reproduction and perpetuation of human ones – conscious or unconscious. As such, diversity biases mostly reflect the diversity bias of AI developers (e.g., in creating the dataset) and HR practitioners (e.g., in the personal characteristics and skills indicated in the job vacancy or the gut feeling they have). In the attempt to identify the most common biases, the respondents of the expert interviews primarily referred to gender and race because of their immediate visibility. Instead, the survey's preliminary results place more emphasis on age and education. Besides, it is worth observing that AI applications can discriminate against a person or a social group by inferring the sensitive information that should not be processed or is legally protected (e.g., race) from other personal and non-personal data. Similarly, technology could target a person or a social group that does not currently enjoy legal protection based on human-comprehensible or incomprehensible personal characteristics.

Because the BIAS project is expected to provide new methods to address diversity bias of AI applications for recruitment and selection purposes, the literature review, the expert interviews, and – to some extent – the mapping exercise examined some potential measures to ensure fairness and respond to diversity bias in the labor market. In addition to the need for legal reform, technical solutions similar to the Debiaser, and the procedural requirements previously mentioned, the expert interviews mentioned the importance of organizational measures. In particular, it is possible to refer to adopting diversity quotas and D&I offices.

In conclusion, all the research outputs will be further examined in this deliverable's final version, due in January 2024. For the time being, ULEID plans to share and discuss them with the rest of the BIAS Consortium, especially to inform the design of the Debiaser (WP3), the performance of the ALTAI (WP3), and the capacity building and raising awareness for the HR and the AI communities (WP5).

9 References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Adams-Prassl, J., Binns, R., & Kelly-Lyth, A. (2023). Directly Discriminatory Algorithms. *The Modern Law Review*, *86*(1), 144–175. https://doi.org/10.1111/1468-2230.12759

Adrián Barragán Díaz, Jimena Y. Ramírez Marín, & Francisco J. Medina Díaz. (2019). The Irony of Choice in Recruitment: When Similarity Turns Recruiters to Other Candidates. *M@n@gement*, *22*(3), 466–486.

AI HLEG. (2019). *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, *7*(2), 205395172094956. https://doi.org/10.1177/2053951720949566

Ajunwa, I. (2020). The “black box” at work. *Big Data & Society*, *7*(2), 205395172096618. https://doi.org/10.1177/2053951720938093

Albert, E. T. (2019). AI in talent acquisition: A review of AI-applications used in recruitment and selection. *Strategic HR Review*, *18*(5), 215–221. https://doi.org/10.1108/SHR-04-2019-0024

Alder, G. S., & Gilbert, J. (2006). Achieving Ethics and Fairness in Hiring: Going Beyond the Law. *Journal of Business Ethics*, *68*(4), 449–464. https://doi.org/10.1007/s10551-006-9039-z

Aloisi, A. (2023). Regulating Algorithmic Management at Work in the European Union: Data Protection, Non-discrimination and Collective Rights. *International Journal of Comparative Labour Law and Industrial Relations*, *40*(1), 1–34.

Aloisi, A., & Potocka-Sionek, N. (2022). De-gigging the labour market? An analysis of the ‘algorithmic management’ provisions in the proposed Platform Work Directive. *Italian Labour Law E-Journal*, 29-50 Pages. https://doi.org/10.6092/ISSN.1561-8048/15027

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Van Vianen, A. E. M., Taris, R., Scholten, E. & Schinkel, S. (2004). Perceived fairness in personnel selection: Determinants and outcomes in different stages of the assessment procedure. *International Journal of Selection and Assessment*, *12*(1), 149–159.

Aristotle, Thomson, J. A. K., Tredennick, H., & Aristotle. (2004). *The Nicomachean ethics* (Further rev. ed). Penguin Books.

Autor, D. H. (2001). Wiring the Labor Market. *Journal of Economic Perspectives*, *15*(1), 25–40. https://doi.org/10.1257/jep.15.1.25

Barocas, S. & Selbst, A. D. (2016). Big Data’s Disparate Impact. *California Law Review*, *104*, 671–732. https://doi.org/10.15779/Z38BG31

Bendick, M., & Nunes, A. P. (2012). Developing the Research Basis for Controlling Bias in Hiring. *Journal of Social Issues*, *68*(2), 238–262. https://doi.org/10.1111/j.1540-4560.2012.01747.x

Funded by the European Union

Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, *94*(4), 991–1013.

Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the People? Opportunities and challenges for participatory AI. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–8. https://doi.org/10.1145/3551624.3555290

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, *63*(2), 215–226. https://doi.org/10.1016/j.bushor.2019.12.001

Bringas Colmenarejo, A., Nannini, L., Rieger, A., Scott, K. M., Zhao, X., Patro, G. K., Kasneci, G., & Kinder-Kurlanda, K. (2022). Fairness in agreement with European Values: An interdisciplinary perspective on AI regulation. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 107–118. https://doi.org/10.1145/3514094.3534158

Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, *8*(1), 205395172098386. https://doi.org/10.1177/2053951720983865

Butucescu, A., & Iliescu, D. (2018). Patterns of change in fairness perceptions during the hiring process: A conceptual replication in a controlled context. *International Journal of Selection and Assessment*, *26*(2–4), 196–201. https://doi.org/10.1111/ijsa.12227

Bygrave, L. A. (2020). Article 22 Automated individual decision-making, including profiling. In Lee A. Bygrave, *The EU General Data Protection Regulation (GDPR)*. Oxford University Press. https://doi.org/10.1093/oso/9780198826491.003.0055

Cappelli, P. (2001). Making the most of on-line recruiting. *Harvard Business Law Review*, *79*(3), 139–146.

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences*, *18*, 13–16. https://doi.org/10.1016/j.cobeha.2017.04.007

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, Feminist Theory and Antiracist Politics. *The University of Chicago Legal Forum*, *1*(8), 139–167.

Collins, P., & Marassi, S. (2021). Is that lawful? Data privacy and fitness trackers in the workplace. *International Journal of Comparative Labour Law and Industrial Relations*, *37*(Issue 1), 65–94. https://doi.org/10.54648/IJCL2021003

Cowgill, B. (2019). *Bias and Productivity in Humans and Machines*. W.E. Upjohn Institute. https://doi.org/10.17848/wp19-309

Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

Dastin, J. (2018, ottobre 11). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Dator, J. (2017). Chapter 3. What Is Fairness? In J. Dator, R. C. Pratt, & Y. Seo (A c. Di), *Fairness, Globalization, and Public Institutions* (pp. 19–34). University of Hawaii Press. https://doi.org/10.1515/9780824841966-004

De Leo, A. & Grossi, T. (2023). *Regulating Platform Work: How This Will Impact Platform Workers?* European Polocy Center and Friedrich-Ebert-Stiftung. https://epc.eu/content/PDF/2023/Regulating_Platform_Work_DP.pdf

Derous, E., & De Fruyt, F. (2016). Developments in Recruitment and Selection Research: Editorial. *International Journal of Selection and Assessment*, *24*(1), 1–3. https://doi.org/10.1111/ijsa.12123

Drosou, M., Jagadish, H. V., Pitoura, E., & Stoyanovich, J. (2017). Diversity in Big Data: A Review. *Big Data*, *5*(2), 73–84. https://doi.org/10.1089/big.2016.0054

Ebert, I., Wildhaber, I., & Adams-Prassl, J. (2021). Big Data in the workplace: Privacy due diligence as a human rights-based approach to employee privacy protection. *Big Data & Society*, *8*(1), 205395172110130. https://doi.org/10.1177/20539517211013051

Equalture. (2023). *Bias Glossary*. https://www.equalture.com/bias-overview/

EU-OSHA. (2023). *Workforce diversity and digital labour platforms: Implications for occupational safety and health*. https://osha.europa.eu/en/publications/workforce-diversity-and-digital-labour-platforms-implications-occupational-safety-and-health

Folger, N., Brosi, P., Stumpf-Wollersheim, J., & Welpe, I. M. (2022). Applicant reactions to digital selection methods: A signaling perspective on innovativeness and procedural justice. *Journal of Business and Psychology*, *37*(4), 735–757. https://doi.org/10.1007/s10869-021-09770-3

Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021). A little bird told me your gender: Gender inferences in social media. *Information Processing & Management*, *58*(3), 102541.

FRA. (2018). *Handbook on European non-discrimination law*. Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-handbook-non-discrimination-law-2018_en.pdf

FRA. (2022). *Bias in algorithms: Artificial intelligence and discrimination*. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

Franck, T. M. (1998). *Fairness in international law and institutions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198267850.001.0001

Fraser, N., & Honneth, A. (2003). *Redistribution or recognition? A political-philosophical exchange*. Verso.

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

Furnham, A., & Chamorro-Premuzic, T. (2010). Consensual beliefs about the fairness and accuracy of selection methods at university: Fairness and accuracy of selection methods at university. *International Journal of Selection and Assessment*, *18*(4), 417–424. https://doi.org/10.1111/j.1468-2389.2010.00523.x

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, *18*(4), 694. https://doi.org/10.2307/258595

Gilliland, S. W., Groth, M., Baker, R. C., Dew, A. E., Polly, L. M., & Langdon, J. C. (2001). Improving applicants' reactions to rejection letters: An application of fairness theory. *Personnel Psychology*, *54*(3), 669–703. https://doi.org/10.1111/j.1744-6570.2001.tb00227.x

Funded by the European Union

Gillis, T. B., & Spiess, J. L. (2019). Big data and discrimination. *The University of Chicago Law Review*, *86*(2), 459–487.

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, *130*, 107179. https://doi.org/10.1016/j.chb.2022.107179

Graham, G. (1988). Two types of feminism. *American Philosophical Quarterly*, *25*(4), 303–312.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, *55*(4). https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2018095.pdf

Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020). Towards a critical race methodology in algorithmic fairness. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 501–512. https://doi.org/10.1145/3351095.3372826

Harned, Z. & Wallach, H. (2020). Stretching human laws to apply to machines: The dangers of a «colorblind» computer. *Florida State University Law Review*, *47*, 617–648.

Harver. (2022). *15 Common Hiring Biases You Need to Avoid*. https://harver.com/resources/e-books/15-common-hiring-biases-you-need-to-avoid/

Hauer, M. P., Kevekordes, J., & Haeri, M. A. (2021). Legal perspective on possible fairness measures – A legal discussion using the example of hiring decisions. *Computer Law & Security Review*, *42*, 105583. https://doi.org/10.1016/j.clsr.2021.105583

Hildebrandt, M. (2020). *Law for computer scientists and other folk* (First edition). Oxford University Press.

Hilliard, A., Guenole, N., & Leutner, F. (2022). Robots are judging me: Perceived fairness of algorithmic recruitment tools. *Frontiers in Psychology*, *13*, 940456. https://doi.org/10.3389/fpsyg.2022.940456

Hossain, S., & Ahmed, S. I. (2021). *Towards a new participatory approach for designing artificial intelligence and data-driven technologies*. https://doi.org/10.48550/ARXIV.2104.04072

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jones, K. (2022). A 'North star' in governing global labour migration? The ILO and the Fair Recruitment Initiative. *Global Social Policy*, *22*(2), 303–322. https://doi.org/10.1177/14680181221084792

Kasirzadeh, A. (2022). Algorithmic fairness and structural Injustice: Insights from feminist political philosophy. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 349–356. https://doi.org/10.1145/3514094.3534188

Kilpatrick, C. & Eklund, H. (2021). Article 21. In Steve Peers, Tamara Hervey, Jeff Kenner & Angela Ward (Eds.), *The EU Charter of Fundamental Rights: A Commentary* (pp. 613–638). Hart Publishing. https://doi.org/10.5040/9781509933495

Kim, P. T. (2019). Big data and artificial intelligence: New challenges for workplace equality. *University of Louiseville Law Review*, *57*(2), 313–328.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, *108*, 22–27. https://doi.org/10.1257/pandp.20181018

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, *10*, 113–174. https://doi.org/10.1093/jla/laz001

Koch-Bayram, I. F., Kaibel, C., Biemann, T., & Triana, M. D. C. (2023). Applicants' experiences with discrimination explain their reactions to algorithms in personnel selection. *International Journal of Selection and Assessment*, *31*(2), 252–266. https://doi.org/10.1111/ijsa.12417

Köchling, A., & Wehner, M. C. (2023). Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses. *International Journal of Selection and Assessment*, *31*(1), 45–62. https://doi.org/10.1111/ijsa.12412

Koivunen, S., Olsson, T., Olshannikova, E., & Lindberg, A. (2019). Understanding decision-making in recruitment: Opportunities and challenges for information technology. *Proceedings of the ACM on Human-Computer Interaction*, *3*(GROUP), 1–22. https://doi.org/10.1145/3361123

Konradt, U., Garbers, Y., Erdogan, B., & Bauer, T. (2016). Patterns of change in fairness perceptions during the hiring process: Patterns of change in fairness. *International Journal of Selection and Assessment*, *24*(3), 246–259. https://doi.org/10.1111/ijsa.12144

Konradt, U., Oldeweme, M., Krys, S., & Otte, K. (2020). A meta-analysis of change in applicants' perceptions of fairness. *International Journal of Selection and Assessment*, *28*(4), 365–382. https://doi.org/10.1111/ijsa.12305

Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply: Fairness in web-based selection. *International Journal of Selection and Assessment*, *21*(2), 155–169. https://doi.org/10.1111/ijsa.12026

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). Towards algorithm auditing: A survey on managing legal, ethical and technological risks of ai, ml and associated algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3778998

Lagioia, F., & Sartor, G. (2020). AI Systems under criminal law: A Legal analysis and a regulatory perspective. *Philosophy & Technology*, *33*(3), 433–465. https://doi.org/10.1007/s13347-019-00362-x

Landon, T. E., & Arvey, R. D. (2007). Ratings of test fairness by human resource professionals. *International Journal of Selection and Assessment*, *15*(2), 185–196. https://doi.org/10.1111/j.1468-2389.2007.00380.x

Langer, M., Baum, K., König, C. J., Hähne, V., Oster, D., & Speith, T. (2021). Spare me the details: How the type of information about automated interviews influences applicant reactions. *International Journal of Selection and Assessment*, *29*(2), 154–169. https://doi.org/10.1111/ijsa.12325

Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, *25*(4), 371–382. https://doi.org/10.1111/ijsa.12191

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, *63*(2), 157–170. https://doi.org/10.1016/j.bushor.2019.10.005

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Malgieri, G. (2020). The concept of fairness in the GDPR: A linguistic and contextual interpretation. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 154–166. https://doi.org/10.1145/3351095.3372868

McKenzie, R. (2018). Bots, bias and big data: Artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Arkansas Law Review*, *71*(2), 529–570.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). *A survey on bias and fairness in machine learning*. https://doi.org/10.48550/ARXIV.1908.09635

Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks. *Digital Society*, *1*(3), 21. https://doi.org/10.1007/s44206-022-00022-2

Mirowska, A., & Mesnet, L. (2022). Preferring the devil you know: Potential applicant reactions to artificial intelligence evaluation of interviews. *Human Resource Management Journal*, *32*(2), 364–383. https://doi.org/10.1111/1748-8583.12393

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4), 44. https://doi.org/10.1007/s11948-021-00319-4

Mucha, H., Robert, S., Breitschwerdt, R., & Fellmann, M. (2020). *Towards participatory design spaces for explainable ai interfaces in expert domains*. https://doi.org/10.24406/PUBLICA-FHG-409866

Mulligan, D. K., Kroll, J. A., Kohli, N., & Wong, R. Y. (2019). This thing called fairness: Disciplinary confusion realizing a value in technology. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–36. https://doi.org/10.1145/3359221

Naim, I., Tanveer, Md. I., Gildea, D., & Hoque, M. E. (2018). Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, *9*(2), 191–204. https://doi.org/10.1109/TAFFC.2016.2614299

Parviainen, H. (2022). Can algorithmic recruitment systems lawfully utilise automated decision-making in the EU? *European Labour Law Journal*, *13*(2), 225–248. https://doi.org/10.1177/20319525221093815

Poulsen, A., Fosch-Villaronga, E., & Søraa, R. A. (2020). Queering machines. *Nature Machine Intelligence*, *2*(3), 152-152.

Rawls, J., & Kelly, E. (2001). *Justice as fairness: A restatement*. Harvard University Press.

Roberson, Q. M. (A c. Di). (2013). *The Oxford handbook of diversity and work*. Oxford University Press.

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to «solve» the problem of discrimination in hiring?: Social, technical and legal perspectives from the UK on automated hiring systems. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 458–468. https://doi.org/10.1145/3351095.3372849

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms*. 64th Annual Meeting of the International Communication Association, Seattle. http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

Schinkel, S., Van Vianen, A., & Van Dierendonck, D. (2013). Selection fairness and outcomes: A field study of interactive effects on applicant reactions: Selection fairness and outcomes. *International Journal of Selection and Assessment*, *21*(1), 22–31. https://doi.org/10.1111/ijsa.12014

Schwarcz, D. & Prince, A. E. (2020). Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Review*, *105*, 1257–1318.

Sharone, O. (2017). LinkedIn or LinkedOut? How social networking sites are reshaping the labor market. In S. Vallas (A c. Di), *Research in the Sociology of Work* (Vol. 30, pp. 1–31). Emerald Publishing Limited. https://doi.org/10.1108/S0277-283320170000030001

Simon, J., Wong, P.-H., & Rieder, G. (2020). Algorithmic bias and the value sensitive design approach. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1534

Sousa, M. J., & Wilks, D. (2018). Sustainable skills for the world of work in the digital age: skills for the digital age. *Systems Research and Behavioral Science*, *35*(4), 399–405. https://doi.org/10.1002/sres.2540

Steiner, D. D., & Gilliland, S. W. (2001). Procedural justice in personnel selection: International and cross-cultural perspectives. *International Journal of Selection and Assessment*, *9*(1 & 2), 124–137. https://doi.org/10.1111/1468-2389.00169

Suranovic, S. M. (2010). *A moderate compromise: Economic policy choice in an era of globalization* (1st ed). Palgrave Macmillan.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: challenges and a path forward. *California Management Review*, *61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Thorsteinson, T. J., & Ryan, A. M. (1997). The effect of selection ratio on perceptions of the fairness of a selection test battery. *International Journal of Selection and Assessment*, *5*(3), 159–168. https://doi.org/10.1111/1468-2389.00056

Tosoni, L. (2021). The right to object to automated individual decisions: Resolving the ambiguity of Article 22(1) of the General Data Protection Regulation. *International Data Privacy Law*, *11*(2), 145–162. https://doi.org/10.1093/idpl/ipaa024

Truxillo, D. M., Bauer, T. N., & Sanchez, R. J. (2001). Multiple dimensions of procedural justice: Longitudinal effects on selection system fairness and test-taking self-efficacy. *International Journal of Selection and Assessment*, *9*(4), 336–349. https://doi.org/10.1111/1468-2389.00185

Truxillo, D. M., Steiner, D. D., & Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment*, *12*(1–2), 39–53. https://doi.org/10.1111/j.0965-075X.2004.00262.x

Van Bekkum, M., & Zuiderveen Borgesius, F. (2023). Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception? *Computer Law & Security Review*, *48*, 105770. https://doi.org/10.1016/j.clsr.2022.105770

Van den Berg, B., & Leenes, R. E. (2013). Abort, retry, fail: Scoping techno-regulation and other techno-effects. In M. Hildebrandt & J. Gaakeer (A c. Di), *Human Law and Computer Law: Comparative Perspectives* (pp. 67–87). Springer Netherlands. https://doi.org/10.1007/978-94-007-6314-2_4

Van Den Bos, K., Vermunt, R., & Wilke, H. A. M. (1997). Procedural and distributive justice: What is fair depends more on what comes first than on what comes next. *Journal of Personality and Social Psychology*, *72*(1), 95–104. https://doi.org/10.1037/0022-3514.72.1.95

Van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, *90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Wachter, S. (2022). *The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law*. https://doi.org/10.48550/ARXIV.2205.01166

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, *41*, 105567. https://doi.org/10.1016/j.clsr.2021.105567

Wesche, J. S., & Sonderegger, A. (2021). Repelled at first sight? Expectations and intentions of job-seekers reading about AI selection in job advertisements. *Computers in Human Behavior*, *125*, 106931. https://doi.org/10.1016/j.chb.2021.106931

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, Solutions, and policy implications. *Journal of Information Policy*, *8*, 78–115. https://doi.org/10.5325/jinfopoli.8.2018.0078

Xenidis, R. (2020). Tuning EU equality law to algorithmic discrimination: Three pathways to resilience. *Maastricht Journal of European and Comparative Law*, *27*(6), 736–758. https://doi.org/10.1177/1023263X20982173

Zibarras, L. D., & Patterson, F. (2015). The role of job relatedness and self-efficacy in applicant perceptions of fairness in a high-stakes selection setting: Selection fairness field study. *International Journal of Selection and Assessment*, *23*(4), 332–344. https://doi.org/10.1111/ijsa.12118

Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, *24*(10), 1572–1593. https://doi.org/10.1080/13642987.2020.1743976

Zytko, D., J. Wisniewski, P., Guha, S., P. S. Baumer, E., & Lee, M. K. (2022). Participatory design of AI systems: Opportunities and challenges across diverse users, relationships, and application domains. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–4. https://doi.org/10.1145/3491101.3516506

Funded by the European Union

10 Annex I: The guidelines of the literature review

Dear Consortium partners,

Because the aim is to investigate the use of AI applications in the labor market, as well as the identification and mitigation of diversity bias of AI arising from them, and to adopt an interdisciplinary approach, ULEID has come up with the following research questions:

1. **What is the present knowledge and application of AI in HR management?**

In this question, we are interested in knowing, context application for instance:

- Recruitment/Selection → Remember that this is our priority because of the future design of the Debiaser!
- Onboarding.
- Assessment/promotion/dismissal.
- Training/development.
- Misconduct/disciplinary processes.
- Management/leadership,
- Churn reduction.
- Burnt out prediction.
- Inclusion.
- Other

But the characteristics of the technology too, such as:

- Level of automation.
- Embodied/non embodied.
- Decision-making power.
- Other.

2. **What is the current, known AI impact on…?**

- HR management (in terms, *e.g.*, better resource utilization, decision-making, problem solving, higher costs).
- Job applicants (in terms, *e.g.*, fairness, discrimination).
- Workers (e.g., proficiency, productivity, fears, distrust, harm, discrimination, higher turnover, upskilling).
- Law and regulation. For instance, are there specific legal obligations exist in comparison with traditional ways of hiring?
- Ethics. For instance, until what extent these processes can be automated? Which ethical requirements are relevant?

3. **Which legal and ethical issues and opportunities arise from the application of AI in HR management?**

- Automated decision-making and human oversight.
- Marginalisation, discrimination, and stigmatisation.
- Fairness and transparency.

Funded by the European Union

- Data protection.
- Robustness.
- Trustworthy AI.

4. **What is bias in AI and, specifically, in AI for HR management?**

This question is more geared towards understanding:

- What does (diversity) bias mean?
- What differences exist between the legal definition of bias and the technical one?
- Which words, practices, sentences express diversity bias?
- To what extent intersectionality is taken into consideration?

5. **How to detect, prevent, or minimize biases in AI and, specifically, in AI for HR management?**

The aim is to mitigate these aspects, especially from a technical angle.

11 Annex II: National Labs Policy

Are you interested in taking part in the BIAS research project and be part of its national labs?

Purpose of the project

The BIAS project is funded by the European Commission under the Grant Agreement no. 101070468 and seeks to identify and mitigate diversity bias that occur when Artificial Intelligence is used in a human resources management context. For this purpose, the BIAS Consortium will develop an innovative technology based on natural language processing (NLP) and case-based reasoning (CBR). More precisely, it will first consult and co-create with you and other stakeholders who have experienced diversity biases in human resources management, have overseen human resources management, and/or have worked on AI applications. In this way, it will be possible to gain a nuanced understanding of what constitutes bias, how bias has been experienced, and how it can be mitigated through both technical and non-technical means. Based on these research findings, the BIAS Consortium will then develop and validate the technology as well as develop capacity building curricula where stakeholders in the AI and HRM community can learn about the existence of AI bias, why it is important to address such bias, and some concrete tools for doing so.

Who is responsible for the research project?

The BIAS project is coordinated by the Norwegian University of Science and Technology (NTNU) (Norway) and is composed of:

- Norwegian University of Science and Technology (NTNU) (Norway);
- Berner Fachhochschule (Switzerland);
- Haskoli Islands (Island);
- Globaz, S.A. (Portugal);
- Crowdhelix Limited (Ireland);
- Smart Venice Srl (Italy);
- Universiteit Leiden (The Netherlands);
- Digiotouch Ou (Estonia);
- Farplas Otomotiv Anonim Sirketi (Türkiye).
- All the partners equally contribute to the implementation of the project activities according to their expertise.

Why are you being asked to participate?

You have been invited to participate in the BIAS project because you have developed, used, experienced, advocated, invested, researched or, more generally, worked on AI applications and/or in the labour market or you represent a minority group. Accordingly, we would like to have you join the creation of the BIAS national lab, namely a community of stakeholders in the field of AI and HR. To this end, we expect each national lab to involve around 100 stakeholders in 7 out of 9 partner's country, namely Estonia, Island, Italy, Norway, Switzerland, The Netherlands, and Türkiye. Stakeholders will be both individuals and organizations acting as:

- Key players, namely AI developers, public and private investors, organisations using AI systems to analyse data;
- Context settlers, namely policy makers, standardisation organisations, and professional networks and platforms for businesses and employees;
- Advocates, namely citizens groups, advocacy organisations, researchers, think tanks, educators, and individual workers.

What does participation involve for you?

If you accept to participate in the BIAS project and, more precisely, its national labs, you will offer your views and expertise on the identification and mitigation of diversity biases in the labour market. The BIAS project is planning many activities that will engage stakeholders in various ways. As a member of the national labs, you will be invited to take part in these activities. You will find out more information about the precise nature of the activities when you are invited to them. How many, and precisely which activities you take part in is up to you. You may also elect to participate in no activities. Some activities we currently have planned are:

- Expert interviews regarding current practices concerning AI in recruitment and HR practices, as well as existing concerns about biases;
- An online survey designed to understand how workers react to and make sense of AI in recruitment and HR practices, as well as existing concerns about biases;
- Co-creation workshops to support BIAS's AI experts in the development of an innovative technology to detected and mitigate diversity biases in AI in recruitment;
- Co-creation workshops for the definition of the BIAS exploitation plan;
- A capacity building programme on biases in AI with an intersectional perspective including online raising awareness activities;
- Qualitative interviews in the frame of an ethnographic study that will be conducted by some of the Consortium partners;
- Professional networking opportunities organized specifically for members of the national labs.

Participation is voluntary

Participation in the project activities is pro-bono and voluntary. If you chose to participate, you can with draw your consent at any time, without giving a reason. All information about you will then be made anonymous. You can choose to be part of the national lab but not participate in any of the activities mentioned above. There will be no negative consequences for you if you choose not to participate or later decide to withdraw.

Your personal privacy – How we will store and use your personal data

We will only use your personal data for the purpose specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation, namely GDPR.

Personal information received will be stored in a secure manner on Microsoft Teams and on the CRM database 'The Raiser's Edge NXT'. Personal information will only be accessible to the Consortium partners and will be password-protected.

Because some of the Consortium partners are located abroad and/or outside of the European Union and the European Economic Area, you should be aware that the data processing is still lawful and secure because of the presence of an adequacy decision between the EU and Switzerland, as well as the signature of a bilateral agreement with our Turkish partner, Farplas Otomotiv Anonim Sirketi.

What will happen to your personal data at the end of the research project?

The BIAS project is scheduled to end on 31st October 2026. Your personal data will be retained for two years after this date because of the audits the European Commission could conduct within this timeframe. The BIAS Consortium will then proceed to the permanent erasure of all the personal data, in the sense that all of them will be made unusable and it will no longer be possible to restore them.

What gives us the right to process your personal data?

We will process your personal data based on your consent.

What rights do I have?

Besides receiving all the information included in this document, you specifically have the right to:

- Withdraw your consent at any time and without negative consequence;
- Obtain from us the confirmations as to whether or not your personal data are being processed. If this is the case, it is possible for you to access them;
- Have your personal data, which are incorrect, incomplete, and/or misleading, rectified;
- Have your personal data been erased;
- Have the processing of your personal data been temporarily restricted;
- Receive all your personal data and eventually have them transmitted to another data controller;
- Object to the processing of your personal data;
- Not to be subject to automated decision-making;
- Lodge a complaint with your national data protection authority;
- Bring a complaint before the competent national court;
- Claim damage for any damage suffered due to unlawful processing of your personal data.

Where can I find out more?

If you have questions about the project, or want to exercise your rights, contact:

NTNU via Project leader Roger A. Søraa, roger.soraa@ntnu.no or Project Administrator Mark Kharas, mark.w.kharas@ntnu.no

Yours sincerely,

Roger A. Søra

Project Leader

12 Annex III: Mapping AI applications in the labor market

Mapping AI applications in the labor market

What is the name of the AI application? ___________

Where did you find it? Insert the link please. ____________

Placing the AI application in context

Where is the AI application deployed?

- Public sector
- Private sector
- Other ________

If possible, could you specify the sector (e.g., academia, think tank, ministry)?

What is the purpose of the AI application?

- Recruitment: Who is the best employer?
- Training: Who and how to improve performance?
- Performance management: Do the working conditions improve job performance?
- Advancement: Who should be promoted?
- Disciplinary measures: Who and how to sanction misconduct in the workplace?
- Retention: Who is likely to leave?
- Other

Is the AI application designed to…?

- Interact with people.
- Make a decision about people (*i.e.*, with a degree of autonomy).
- Help a human being (*e.g.*, a recruiter) make a decision about people (*e.g.*, the candidate).
- Other

Is the applicant or employee aware of the involvement of the AI in the application process?

- Yes
- No
- I do not know

If you answered 'yes' in the previous question, how is the person informed about the involvement of AI in the application process?

- The system is transparent and explainable.
- There is information available to the candidate (*e.g.*, in the candidature form).
- The company sends a consent form beforehand.
- Other ________

Is the AI application…?

- Self-learning.

- Overseen by a human-in-the-loop (i.e., human intervention in each decision cycle of the system).
- Overseen by a human-on-the-loop (i.e., human intervention during the design cycle of the system and monitoring the system's operation).
- Overseen by a human-in-command (i.e., the human being can decide when and how to use the AI application in any particular situation).
- I do not know.
- Other ______

Examining respect for diversity in the AI application

Which kind of information does the AI application process?

- Gender
- Racial and/or ethnic origin
- Nationality
- Age
- Disability
- Sexual orientation
- Family status
- Address
- Health
- Genetics
- Education
- Previous working experience
- Gap years
- Religion
- Trade union affiliation
- Political affiliation
- Language(s) spoken
- Other _______

The developers of the AI system may have defined what a desirable employee is according to the company policies. Is this information available? How does the company define 'desirable employee'? ___________

Are there in place measures to guarantee that data used to develop the AI application are up-to-date, complete, and representative of the environment they will be deployed in?

- Yes
- No
- I do not know

Is there a strategy or a set of procedures to avoid creating or reinforcing unfair bias?

- Yes
- No
- I do not know

If there is a strategy or a set of procedures to avoid creating or reinforcing unfair bias, which of the following did they implement?

- Conducted an impact assessment.
- Updated the code of conduct.
- Carried out the Trustworthy AI Assessment List.
- Invited different stakeholders to provide feedback on the tool.
- Other______

Has the company considered diversity and representativeness of end-users and/or subjects in the data?

- Yes
- No
- I do not know

If yes, which users and how were they involved? _________

Is there a mechanism that allows for the flagging of issues related to bias, discrimination, or poor performance of the AI application?

- Yes
- No
- I do not know

If yes, what issues have been flagged? __________

Is the AI application usable by those with special needs or disabilities or those at risk of exclusion?

- Yes
- No
- I do not know

If yes, how and what characteristics were taken in account for it? _________

Do you consider the AI application fair because…?

- It embodies an ideal of equal treatment between individuals (equality).
- It embodies an ideal of providing the support each person needs (equity).
- It embodies an ideal of removing systematic barriers (justice).
- I could not say.
- Other ______

13 Annex IV: Mapping some AI applications

Name	Website
Alight	https://www.alight.com/it
AssesFirst	https://www.assessfirst.com/en/
AvrioAI	https://www.avrioai.com
BrainFirst	https://www.brainsfirst.com
ChatGPT	https://openai.com/gpt-4
CVVIZ	https://cvviz.com
Fetcher	https://fetcher.ai
HiJob	https://www.hijob.me/en
Harver	https://harver.com
HireVue	https://www.hirevue.com
Ideal	https://ideal.com
Inda	https://inda.ai/en/
InRecruting	https://www.in-recruiting.com/it/
InTouch	https://keepmeintouch.io
Jacando	https://www.jacando.com/en/
Joboti	https://www.joboti.com/en/
JobPal	https://jobpal.ai/en/
Lattice	https://lattice.com
Manatal	https://www.manatal.com
MeetFrank	https://www.meetfrank.com/business
MeMotive	https://www.memotive.app/EN/index.html
Namely	https://www.namely.com
Oracle AI Apps for HR	https://www.oracle.com/human-capital-management/ai-at-work/ai/
Pusula360	https://www.pusula360.com
RecruitRobin	https://www.recruitrobin.com/en/
SAP	https://www.sap.com/italy/products/artificial-intelligence/recruit-to-retire.html
Solique	https://www.solique.ch/english/
Stepstone	https://www.thestepstonegroup.com/en/solutions/recruiter/
Talentia	https://www.talentia-software.com/it/
Talview	https://www.talview.com
Textkernel	https://www.textkernel.com
Turing	https://www.turing.com/hire-developers
Validata	https://validatagroup.com
WittyWorks	https://www.witty.works
Wonderkind	https://wonderkind.com
Workable	https://www.workable.com/recruiting-software
Workday	https://www.workday.com/it-it/solutions/role/enterprise-hr/overview.html
Yenibiris	https://www.yenibiris.com
Zoho	https://www.zoho.com

14 Annex V: The template of the expert interviews

Gender (M/F/other):	
Constituent type:	**AI Developers**
Interviewer name:	
Date of interview:	

General instructions for the interviewer

• Duration of the interview: 45 minutes
• Fill in one form per interview. Make a copy of the tab "template".
• Process of the interview: Either you fill in the form during the interview (optionally with the help of a colleague), or you can record the interview and fill in the form afterwards. **We recommend having a colleague help during the interview. Voice recording requires consent and may have implications with your national data protection laws**
• Do not read the list of options to the interviewee, let them bring their own ideas first!
• For any questions about the expert interviews, your contact in BIAS: Linda Rafnsdóttir, glr@hi.is. If your interview subjects have questions about the BIAS project in general, they should contact Mark Kharas, mark.w.kharas@ntnu.no

Introductory speech

*Instruction: Deliver this speech at the beginning of the interview. You can adapt the text, but the elements in **bold** are the ones that must be compulsorily conveyed.*

In this European project, BIAS, the main objective is to investigate and mitigate possible biases that appear in Artificial Intelligence (AI). Our research project is specifically concerned with the use of AI in the employment sector to manage and control individual workers. This can involve analyzing text created by an employee or recruitment candidate in order to assist management in deciding to invite a candidate for an interview, to training and employee engagement, or to monitor for infractions that could lead to disciplinary proceedings. However, we know that AI is used in many other areas where the issue of bias is also important. We are therefore asking you questions about AI in general in addition to AI specifically used in a human resource management context.

In this interview, we would like to get your view on this development, and whether you believe, that this can lead to biased decisions that run contrary to the goals of social rights and fairness in relationship to work and employment.

Two of the main goals of BIAS is to develop new bias identifying and bias preventing AI algorithms and developing training materials for AI developers and human resource professionals concerning bias in AI. Your answers during this interview will directly contribute to these two objectives

RGPD notice - use of data

Instruction: You must explain that the interview will be anonymous.

The participation to this interview is fully anonymous. Your identity will never be linked to the responses you provide. Your name and contact information will be stored by the project partners only for organizational purposes of the interviews during the project.

RGPD notice - recording *(only if recording)*

Instruction: Only if you choose to record the interview - Inform the interviewee of the recording and its objectives, and ask for their explicit consent.

I am going to record the interview to facilitate my note taking. This recording will only be stored on my computer and deleted immediately after I have processed your answers. Do you allow this?

Answer:

Yes	
No	

SECTION 1: GENERAL QUESTIONS

1 **What type of organization do you represent?**
Instruction: *put a "X" next to each relevant item. Fill in the "Other" slot if needed.*

Answer:

Public research organization/university	
Private research organization/university	
Company/corporation	

How many employees does your organization have?

Other:

2 **Which kind of bias – if any - are in your opinion more likely to be reproduced/strengthened through AI solutions? Please provide some examples**

Answer:

3 **When developing an AI solution, biases might occur at different levels: data that is inputted into the model, algorithms performed on the data, and outcomes of the models. Which of those bias do you think are problematic/difficult to tackle and why/how?**
Instruction: *put a "X" next to each relevant item. Fill in the "Other" slot if needed.*

Answer:

Data	
Algorithms	
Output	

Explination:

4 **Did you attend any course about ethics in AI within your university curricula or beyond it? which one(s)?**
Instruction: *put a "X" next to each relevant item. Fill in the "Other" slot if needed.*

Answer:

Yes	
No	

Explination:

SECTION 1: RECRUITMENT

5 To what extent is AI (algorithms) technology being produced for use in recruitment and selection?
Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

In this country
Elsewhere

Comments

6 Do you think that the use of AI (algorithms) is in general positive in recruitment and selection?
Instruction: *Answer the question and ask for explination for yes or no answer; make notes in the notes field*

Answer:
Yes/Sometimes
No

If yes: Why and when is it positive?
If no: Why is it not positive

7 Do you think there is a risk that AI (algorithms) will introduce bias (exclusion or discrimination) in recruitment/selection decision making?
Instruction: Answer the question and ask for explination for yes or no answer; make notes in the notes field

Answer:
Yes
No

If yes—what kind of bias?
If no – why not?

8 Does your organization use AI (algorithms) in recruitment and selection?
Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:
Yes
No
I don't know

If yes: how much do you think your company uses it compared to other similar organizations?
Answer:
We use it more
We use it about the same
We use it less

If no: Why don't you use AI in recruitment and selection?
If yes: Why do you use it similarly or differently than other organizations?

9 What sort of data about individuals do you generally gather for use in the AI algorithms you develop
Instruction: Answer from the options provided. Enter any comments/notes from the conversation

General comments

SECTION 2: MANAGEMENT

10 **To what extent is technology being produced with the purpose of using AI (algorithms) to manage and control individual employees?**
Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

In this country
Elsewhere

Comments

11 **Do you think that using AI (algorithms) to manage and control individual employees is, in general, a good practice?**
Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it positive and why?
If no, why not?

If they answered yes to question 11, ask questions 11a, 11b, and 11c; otherwise move to question 12

11a **Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to productivity?**
Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it acceptable and why?
If no, why not?

11b **Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to infractions of companies rules and policies?**
Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it acceptable and why?
If no, why not?

11c **Are there other areas where it is acceptable to use AI to manage and control individual employees**

General comments

12 **Do you think there is a risk that AI will introduce bias (exclusion or discrimination) in decision making?**
Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes
No

If yes, what kind of bias?
If no, why not?

13 **Do you think that AI developers take action to screen their algorithms for 'fairness'?**

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:

Yes
No

If yes, what kind of action?
If no, why not?

14 **Does your organization use AI (algorithms) in to manage and control employees?**

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:

Yes
No

If yes: how much do you think your company uses it compared to other similar organizations?

Answer:

We use it more
We use it about the same
We use it less

If no: Why don't you use AI to manage and control employees?
If yes: Why do you use it similarly or differently than other organizations?

15 **If you were to receive training concerning bias in AI, what topics would you most want addressed**

Answer:

16 **Do you accept being invited to the project's future events or contacted for other research activities related to BIAS? We will not use your contact information for any other purpose.**

Gender (M/F/other):	
Constituent type:	**HR Executive**
Interviewer name:	
Date of interview:	

General instructions for the interviewer

- Duration of the interview: 45 minutes
- Fill in one form per interview. Make a copy of the tab "template".
- Process of the interview: Either you fill in the form during the interview (optionally with the help of a colleague), or you can record the interview and fill in the form afterwards. **We recommend having a colleague help during the interview. Voice recording requires consent and may have implications with your national data protection laws**
- Do not read the list of options to the interviewee, let them bring their own ideas first!
- For any questions, your contact in BIAS:

Introductory speech

Instruction: Deliver this speech at the beginning of the interview. You can adapt the text, but the elements in ***bold*** *are the ones that must be compulsorily conveyed.*

In this European project, BIAS, the main objective is to investigate and mitigate possible biases when Artificial Intelligence (AI) is used in the employment sector to manage and control individual workers. This can involve analyzing text created by an employee or recruitment candidate in order to assist management in deciding to invite a candidate for an interview, to training and employee engagement, or to monitor for infractions that could lead to disciplinary proceedings.

In this interview, we like to get your view on this development, and whether you believe, that this can lead to biased decisions that run contrary to the goals of social rights and fairness in relationship to work and employment.

RGPD notice - use of data

Instruction: You must explain that the interview will be anonymous.

The participation to this interview is fully anonymous. Your identity will never be linked to the responses you provide. Your name and contact information will be stored by the project partners only for organizational purposes of the interviews during the project.

RGPD notice - recording *(only if recording)*

Instruction: Only if you choose to record the interview - Inform the interviewee of the recording and its objectives, and ask for their explicit consent.

I am going to record the interview to facilitate my note taking. This recording will only be stored on my computer and deleted immediately after I have processed your answers. Do you allow this?

Answer:

Yes	
No	

1 **What type of organization do you represent?**

Instruction: *put a "X" next to each relevant item. Fill in the "Other" slot if needed.*

Answer:

Public research organization/university	
Private research organization/university	
Company/corporation	

How many employees does your organization have?

Other:

SECTION 1: RECRUITMENT

2 **1.To what extent do believe that employers or recruitment agencies use AI (algorithms) in recruitment and selection?**

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

In this country ☐
Elsewhere ☐

Comments

---

3 **Do you think that the use of AI (algorithms) is in general positive in recruitment and selection?**

Instruction: *Answer the question and ask for explination for yes or no answer; make notes in the notes field*

Answer:

Yes/Sometimes ☐
No ☐

If yes: Why and when is it positive?
If no: Why is it not positive

---

4 **Do you think there is a risk that AI (algorithms) will introduce bias (exclusion or discrimination) in recruitment/selection decision making?**

Instruction: Answer the question and ask for explination for yes or no answer; make notes in the notes field

Answer:

Yes ☐
No ☐

If yes—what kind of bias?
If no – why not?

If they answered yes to question 4, ask question 4a, otherwise proceed on to question 5

4a **Do you know of any strategies to mitigate such bias?**

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:

Yes ☐
No ☐
I don't know ☐

If yes: What are the strategies

---

5 **Does your organization use AI (algorithms) in recruitment and selection?**

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:

Yes ☐
No ☐
I don't know ☐

If yes: how much do you think your company uses it compared to other similar organizations?

Answer:

We use it more ☐
We use it about the same ☐
We use it less ☐

If no: Why don't you use AI in recruitment and selection?
If yes: Why do you use it similarly or differently than other organizations?

If they answered yes to BOTH questions 4 AND, ask question 45a, otherwise proceed on to question 6

5a **Does your organization use AI any of the mitigation ideas you mentioned eaerlier?**

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:

Yes ☐
No ☐
I don't know ☐

If yes: Do you think these mitigation efforts are successful?

Answer:

Yes ☐
No ☐
I don't know ☐

If no: Why not?
If yes: Which ones?

---

6 **What sort of data about individuals do you generally gather during the recruitment process?**

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

General comments

SECTION 2: MANAGEMENT

7 To what extent do believe that employers or recruitment agencies use AI (algorithms) to manage and control individual employees

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

In this country
Elsewhere

Comments

8 Do you think that using AI (algorithms) to manage and control individual employees is, in general, a good practice?

Instruction: Answer from the options provided. Enter

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it positive and why?
If no, why not?

If they answered yes to question 8, ask questions 8a, 8b, and 8c; otherwise move to question 9

8a Do you think that AI is in general acceptable to manage and control individual workers by monitoring activities due to disciplinary motives related to productivity?

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it acceptable and why?
If no, why not?

8b Do you think that AI is in general acceptable to manage

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:
Yes/Sometimes
No

If yes/sometimes, when is it acceptable and why?
If no, why not?

8c Are there other areas where it is acceptable to use AI to manage and control individual employees

General comments

9 Do you think there is a risk that AI will introduce bias (exclusion or discrimination) in decision making?

Instruction: Answer from the options provided. Enter

Answer:
Yes
No

If yes, what kind of bias?
If no, why not?

10 Do you think that AI developers take action to screen their algorithms for 'fairness'?

Instruction: Answer from the options provided. Enter any comments/notes from the conversation

Answer:

Yes [ ]
No [ ]

If yes, what kind of action?
If no, why not?

[ ]

11 Does your organization use AI (algorithms) in to manage and control employees?

Instruction: *Answer from the options provided. Enter any comments/notes from the conversation*

Answer:

Yes [ ]
No [ ]

If yes: how much do you think your company uses it compared to other similar organizations?

Answer:

We use it more [ ]
We use it about the same [ ]
We use it less [ ]

If no: Why don't you use AI to manage and control employees?
If yes: Why do you use it similarly or differently than other organizations?

[ ]

12 If you were to receive training concerning bias in the use of AI in human resource management, what topics would you most want addressed?

Answer:

[ ]

13 Do you accept being invited to the project's future events or contacted for other research activities related to BIAS? We will not use your contact information for any other purpose.

15 Annex VI: The survey

Introduction to the survey

Thank you for participating in our survey on discrimination, exclusion, and marginalization of workers that the use of artificial intelligence (AI) applications in the labor market causes sometimes.

This survey is part of the BIAS project - Mitigating diversity biases of AI in the labor market, which is funded by the European Commission (Grant agreement no. 101070468) and the Swiss State Secretariat for Education, Research and Innovation.

Overall, **the BIAS project seeks to identify and address some prejudices and other unsupported judgements against one worker because of their personal characteristics in a way considered unfair.**

We are therefore **asking you questions about your experience and attitudes towards AI applications in the labor market.**

The survey is divided in four sections:

1. **Interaction**. We seek to understand **whether you have ever interacted with one or more AI applications in the labor market and, if so, how it worked.** For example, LinkedIn provides algorithms to advertise job vacancies and actively search and approach job candidates. At the same time, some large companies use their algorithms to hire, manage, or fire people.
2. **Experience.** We then move to your personal experience. **In other words, how has the use of AI applications to hire, manage, or fire made you feel?**
3. **Perception**. Also, we would like **to hear your general attitude toward the use of AI applications in the labor market.** Is it something we should pursue or not?
4. Lastly, we ask for **some personal information**. We know that you might perceive some questions (*e.g.*, sexual orientation and disability) as sensitive and intimate. But, **because the BIAS project focuses on discrimination, exclusion, and marginalization, it is important for us to understand who and how diverse our respondents are. We aim to be as inclusive as possible**. Please, remember that it will not be possible to determine your identity using this information and you can always say that you prefer not to respond to a certain question.

It should approximately take you 10 minutes to fill out the survey.

Please answer all questions. Your honest opinion is greatly appreciated, and your information will be anonymized.

If you would like to know more about the BIAS project, you can access our website.

By answering 'yes' to the question below, you acknowledge that your participation in this research is voluntary, that you are at least 18 years of age, and that you have the right to withdraw at any point during the study, for any reason, and without any prejudice.

The survey is fully anonymous and targets every worker who has interacted with AI applications in the labor market in Europe. The results of the survey will inform a report to be produced by the BIAS Consortium to the European Commission by July 2023. Eventually, they will also be disseminated through academic conferences and publications. In any case, no personally identifying information will be included.

If you have any questions and/or you would like to be involved in other research activities, you can reach the research team at info@biasproject.eu.

Please be aware that all the versions of the survey other than the English one have been translated by the members of the BIAS Consortium or external volunteers. We apologize for any mistake.

Yours sincerely,

Carlotta Rigotti & Eduard Fosch-Villaronga

eLaw Center for Law and Digital Technologies

Leiden University, the Netherlands.

- Yes, I would like to participate.
- No, I would not like to participate.

Preliminary questions about the AI applications you interact or have interacted with.

Q.1 Have you ever interacted with one or more AI applications in the labor market?

- Yes
- No
- Maybe

Q.1.1 Do you know that LinkedIn and any other social media platform for professional networking and career development use AI?

- Yes
- No

Q.1.2 Do you know that employers increasingly use AI applications to identify you, screen your job application, hold job interviews, train their employers, improve teamwork, and decide to promote or fire someone?

- Yes
- No

Q.1.3 Based on your previous answers, would you change your reply to the question "Have you ever interacted with one or more AI applications in the labor market?"

- Yes
- No
- Maybe

Q.2 In which sector do you work in?

- The public sector
- The private sector
- Other

Q.3 If possible, could you please specify the sector? Multiple answers are possible.

- Agriculture, fisheries, forestry
- Mineral extraction
- Manufacturing
- Energy supply
- Water utilities
- Waste management
- Construction
- Wholesale and retail
- Transport and storage
- Healthcare
- Social welfare
- Financial services
- Insurance services
- IT
- Real estate rental and trade
- Specialist business services
- Education
- Culture
- Sport
- Food
- Accommodation
- Other

Q.4 Are you…?

- Employed, on a permanent basis.
- Employed, on a temporary basis.
- Self-employed
- A trainee
- Other

Q.5 How would you describe the AI applications you have interacted with? Multiple answers are possible.

- It put me in contact with possible employers.
- It analyzed some personal information of mine (*e.g.*, in the CV, motivation letter).
- It replaced a human being in a job interview.
- It tested my skills, for example, by inviting me to play a game.
- It trained my skills.
- It gave me information about my tasks, shifts, and other practicalities.
- It checked my time availability.
- It monitored my productivity.

- It ranked me.
- It checked my health conditions.
- Other

Q.6 What were the AI applications you have interacted with used for? Multiple answers are possible.

- To recruit you.
- To train you.
- To improve your performance at work.
- To decide to promote you.
- To sanction your possible misconduct.
- To foresee your possible conduct (*e.g.*, health problems, dismissal).
- Other

Q.7 When interacting with AI applications, have you been informed about its use?

- Yes
- No
- I do not remember.

Q.7.1 If you have been informed about the use of AI applications, when were you informed about?

- Before interacting with the AI applications
- After interacting with the AI applications
- Before and after interacting with the AI applications
- I do not remember.

Q.7.2 If you have been informed about the use of AI applications, what were you informed about? Multiple answers are possible.

- How your personal data were expected to be used.
- How the AI application worked.
- Why the AI application was used.
- The possibility to get further clarification about the AI application.
- The possibility to object to the data processing being automatically done by the AI application, without the involvement of a human being.
- Other

Q.7.3 If you have been informed about the use of AI applications, how was the information presented to you?

- Orally
- Written
- Orally and written.
- I do not remember.

Q.8 Could you opt out of the use of AI applications?

- Yes

- No
- Sometimes
- I do not remember.

Q.9 Were the AI applications used to…? Multiple answers are possible:

- Interact with you.
- Make a decision about you.
- Help another person (*e.g.*, your boss) make a decision about you.
- Other

Q.10 Do you know if…?

- The AI applications were fully automated and there was no human intervention.
- A person monitored each decision made by the AI application.
- A person decided whether, when, and how to use the AI on a case-by-case basis.
- I do not know whether there was a human involved in the process.

Your personal attitudes towards the use of AI applications in the labor market

Q.11 Have you felt comfortable with the use of AI applications in the labor market?

- Very comfortable
- Slightly comfortable
- Neither comfortable, nor uncomfortable
- Uncomfortable
- Very uncomfortable
- Other: ______
- Prefer not to respond.

Q.12 When interacting with AI applications, have you ever been directly asked about your…? Multiple answers are possible:

- Gender
- Racial and/or ethnic origin
- Nationality
- Age
- Disability
- Sexual orientation
- Family status
- Address
- Physical health
- Mental health
- Genetics
- Education
- Previous working experience
- Gap years
- Religion

- Trade union affiliation
- Political affiliation
- Language(s) spoken
- Availability
- Other: _______

Q.13 Do you believe that the personal information in the previous question could be used to negatively discriminate against you by AI applications? Negative discrimination means you have been disadvantaged because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure.

Q.14 Do you think some other information (*e.g.*, your hobbies, domestic routine) were asked to negatively discriminate against you? Negative discrimination means you have been disadvantaged because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure.

Q.14.1 If yes, could you explain further? _______

Q.15 Do you believe that the personal information in the previous question could be used to positively discriminate against you by AI applications? Positive discrimination means you have been favored because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure.

Q.16 Do you think some other information (*e.g.*, your hobbies, domestic routine) were asked to positively discriminate against you? Positive discrimination means you have been favored because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure.

Q.16.1 If yes, could you explain further? _______

Q.17 Do you think that the collection of the said personal information could be important to later adequately perform the job tasks?

- Yes
- No
- I am not sure.

Q.18 Do you feel that you experienced negative discrimination from the disclosure of this information? Negative discrimination means you have been disadvantaged because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure.

Q.18.1 If yes, could you explain further? ______

Q.19 Do you feel that you have experienced positive discrimination from the disclosure of this information? Positive discrimination means you have been favored because of one or more personal characteristics (*e.g.*, gender, age, disability).

- Yes
- No
- I am not sure

Q.19.1 If yes, could you explain further? ______

General public attitudes towards the use of AI applications in the labor market

Q.20 Generally speaking, which view do you have of the use of AI applications in the labor market?

- Very positive
- Fairly positive
- Neither positive, nor negative
- Fairly negative
- Very negative
- I do not know.
- Other

Q.21 Using a scale from 1 to 7, please tell me what impact AI applications might have on the labor market. '1' means that you consider the impact very positive and '7' means that you consider it very negative.

- The personal autonomy of the worker
- Social inclusion
- The transparency of any decision-making
- Data protection
- The accountability of the employer
- The mental well-being of the worker
- Productivity

Q.22 Do you think there is any other positive or negative impact that was not mentioned in the previous question? ____________

Q.23 In your opinion, when could AI applications be defined fair in the labor market? Multiple answers are possible.

- When it does not discriminate against one or more people because of their personal characteristics (*e.g.,* gender, age, disability).
- When it favors one or more people because of their personal characteristics (e.g, gender, age, disability).
- When it treats similar people (*e.g.*, all Black women) in a similar way.
- When it can perform better than a human being working in human resources.
- When the worker knows how the AI application works.
- When it is possible to explain how the AI application has made a certain decision.
- When it is possible for the worker to challenge the decision made by the AI application.
- When a human being monitors how the AI application works and reviews its decisions.
- When the AI application allows the individual to work in a way that suits them.
- There will never be a fair AI application in the labor market.
- Other

Demographic questions

Q.24 What country do you work in? Multiple answers are possible.

- Austria
- Belgium
- Bulgaria
- Croatia
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Latvia
- Lithuania
- Luxemburg
- Malta
- The Netherlands
- Norway
- Poland
- Portugal
- Republic of Cyprus
- Romania
- Slovakia
- Slovenia
- Spain

- Sweden
- Switzerland
- Türkiye
- Other
- Prefer not to respond.

Q.25 What is your nationality? Multiple answers are possible.

- Austrian
- Belgian
- British
- Bulgarian
- Croatian
- Cypriot
- Czech
- Danish
- Dutch
- Estonian
- Finnish
- French
- German
- Greek
- Hungarian
- Icelandic
- Irish
- Italian
- Latvian
- Lithuanian
- Luxemburg
- Maltese
- Norwegian
- Polish
- Portuguese
- Romanian
- Slovakian
- Slovenian
- Spanish
- Swedish
- Swiss
- Turkish
- Other
- Prefer not to respond.

Q.26 Please, indicate your age group:

- 18-28
- 29-39

- 40-50
- 51-61
- Over 61
- Prefer not to respond.

Q.27 What is your highest level of completed education?

- Primary school
- Secondary school
- Vocational training
- Bachelor's degree
- Master's degree
- PhD
- Other
- Prefer not to respond.

Q.28 What gender do you self-identify as? Multiple answers are possible.

- Female
- Male
- Transgender
- Non-binary
- Other
- Prefer not to respond.

Q.29 What sexual orientation do you self-identify as?

- Straight
- Lesbian
- Gay
- Bisexual
- Asexual
- Other
- Prefer not respond.

Q.30 Which of the following designations do you self-identify as? Multiple answers are possible.

- Black
- Latin American
- White
- Asian
- Middle Eastern
- North African
- Roma
- Multiple racial/ethnic groups
- Other
- Prefer not to respond.

Q.31 Considering the country you live in, do you consider that the previous designation(s) make you part of a minority?

- Yes
- No
- Prefer not to respond.

Q.32 Do you self-identify as a person with physical disability?

- Yes
- No
- Prefer not to respond.

Q.33 Do you self-identify as a person with mental disability?

- Yes
- No
- Prefer not to respond.

Q.34 Are you affiliated to a trade union?

- Yes
- No
- Prefer not to respond.

Q.35 Do you consider yourself to…?

- Be affiliated to a political party because you hold membership.
- Be affiliated to a political party because you support it and vote for it.
- Have strong political opinions.
- Have publicly shared your political opinions (*e.g.*, via social media, public demonstrations).
- I am not affiliated to a political party, nor I have political opinions.
- Other
- Prefer not to respond.

Q.36 Do you consider yourself to be…?

- Catholic
- Orthodox Christian
- Protestant
- Jewish
- Muslim - Shia
- Muslim - Sunni
- Sikh
- Buddhist
- Hindu
- Atheist
- Agnostic
- Other
- Prefer not to respond.

Thank you very much for your participation in the BIAS survey!

If you have any further questions and/or you would like to be involved in other research activities, you can reach the research team at info@biasproject.eu.

BIAS

Mitigating biases
of AI in the
labour market

Consortium

NTNU

UNIVERSITY OF ICELAND

LOBA

Crowdhelix

SMARTVENICE

Universiteit Leiden eLaw

Digiotouch

farplas

Bern University of Applied Sciences

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the Swiss State Secretariat for Education, Research and Innovation can be held responsible for them.

email
info@biasproject.eu

website
www.biasproject.eu

social media
@BIASProjectEU

The Associated Partner Bern University of Applied Sciences has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI)