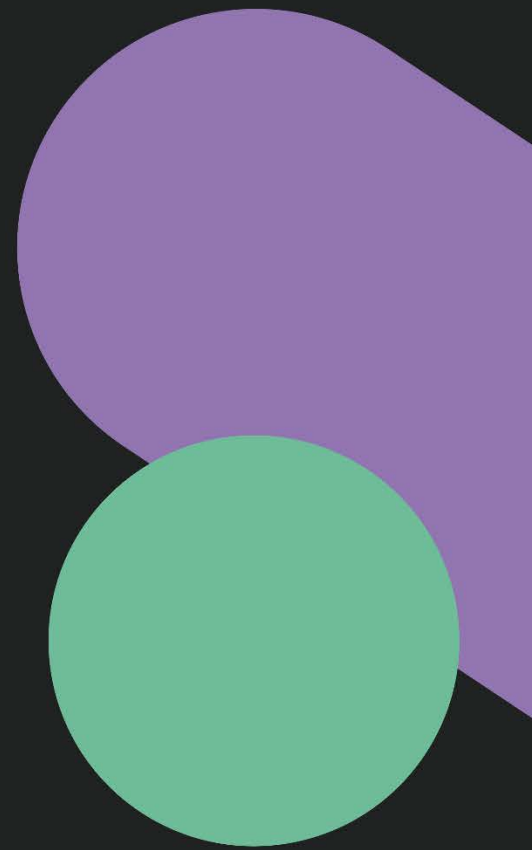
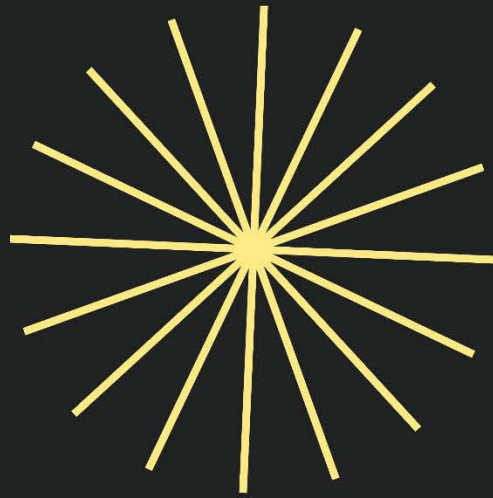


BIAS

Mitigating biases
of AI in the
labour market



D2.4

Final report on co-creation methodologies and findings

Contact us

info@biasproject.eu

www.biasproject.eu



LOBA[®]

CrowdHelix

 Funded by
the European Union



farplas



@BIASProjectEU





D2.4 Final report on co-creation methodologies and findings

Deliverable type	Month and date of delivery
Report	January, 2024
Work package	WP Lead Beneficiary
2	ULEID
Dissemination level	Authors
Public	Marzia Cescon, Simona Lo Giudice, Maria Sangiuliano

Programme	Contract Number	Duration	Start
Horizon Europe	101070468	48 months	November 2022





Main Authors

Name

Marzia Cescon
Simona Lo Giudice
Maria Sangiuliano

Organisation

SVEN
SVEN
SVEN

Contributors

Name

Mascha Kurpicz-Briki
Alex Puttick
Pinar Øzturk

Organisation

BFH
BFH
NTNU

Peer Reviews

Name

Öznur Karakaş
Carlotta Rigotti

Organisation

NTNU
ULEID

Revision History

Version	Date	Reviewer	Modifications
0.1	15.01.2024	NTNU, ULEID	
1.0 Final	30.01.2024	SVEN	

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.





Abbreviation	Meaning
AI	Artificial Intelligence
AI HLEG	High-Level Expert Group on Artificial Intelligence
ALTAI	Assessment List of Trustworthy Artificial Intelligence
API	Application programming interface
ATS	Applicant Tracking System
BS	Bachelor of Science
CBR	Case Based Reasoning
ChatGPT	Chat Generative Pre-trained Transformer
CSO	Civil Society Organization
CV	Curriculum Vitae
DSS	Decision Support System
EDI	Equality, Diversity and Inclusion
EIBD	Emergent Intersectional Bias Detection
GA	Grant Agreement
GE	Gender Equality
HR	Human Resources
IAT	Implicit Association Test
IBD	Intersectional Bias Detection
IT	Information Technology
KPI	Key Performance Indicator
LGBTIQA+	Lesbian, Gay, Bisexual, Transgender, Intersex, Queer/Questioning, Asexual
ML	Machine Learning
NGO	Non-Governmental Organization
NLP	Natural Language Processing
SDK	Software Development Kit
SMEs	Small and Medium Enterprises
WEAT	Work Embedding Association Test
WP	Working package





1. Table of contents

- 1. Table of contents5**
- 2. Table of figures.....8**
- 3. Table of Tables9**
- 4. Executive Summary 10**
- 5. Introduction and methodological approach 11**
- 6. Cross-cutting methodological aspects for the two workshops on engagement and preparation 16**
 - 6.1 Targeted participants/stakeholders 16
 - 6.2 Preparatory and engagement activities..... 17
- 7. The first workshop’s specific co-creation methodology 19**
 - 7.1 Workshop’s agenda and target 19
 - 7.2 Introduction & BIAS presentation 20
 - 7.3 Panel discussion..... 20
 - 7.4 Co-creation group work 22
 - 7.5 Reporting process 27
- 8. Implementation of the first co-creation workshops: results..... 30**
 - 8.1 Overview of the conducted workshops 30
 - 8.2 Overall implementation of the workshops 31
 - 8.3 Summary of the reports 34
 - 8.3.1 1st co-creation workshop in Italy 34
 - 8.3.2 1st co-creation workshop in Norway 39
 - 8.3.3 1st co-creation workshop in The Netherlands..... 42
 - 8.3.4 1st co-creation workshop in Iceland..... 46
 - 8.3.5 1st co-creation workshop in Turkey 49
 - 8.3.6 1st co-creation workshop in Estonia..... 53
 - 8.3.7 1st co-creation workshop in Switzerland..... 57
 - 8.4 Workshops’ core outputs: the wordlists..... 61
 - 8.4.1 Overview of the collected wordlists 61
 - 8.4.2 Use of the wordlists in the BIAS technological development 64
- 9. Methodology of the second co-creation workshop..... 66**
 - 9.1 Workshop’s agenda and target 66
 - 9.2 Introduction & BIAS presentation 67
 - 9.3 Discussion in two groups: how does a fair HR recruitment process look like?..... 67
 - 9.4 Interactive/hands on work: which requirements for AI tools in recruiting? 70
 - 9.5 Reporting process 76





10. Implementation of the second co-creation workshops: results 77

10.1 Overview of the conducted workshops 77

10.2 Discussion into groups: what does a fair HR recruitment process look like? 78

10.2.1 Activity one: ranking principles – Results 78

10.2.2 Activity two: Objectivity and Consistency - Findings 80

10.2.3 Discussions into groups: takeaways..... 87

10.3 Interactive/hands on work: which requirements for AI tools in recruiting? 88

10.3.1 Feedback from the rapporteurs..... 88

10.3.2 The perceived needs and requirements for AI tools 88

10.3.3 Interactive work - Challenges and Takeaways 96

11. The International Co-Creation Workshop- Methodology 98

11.1 Workshop’s agenda and target 98

11.2 First Block - Simulating the use of AI based systems in Recruitment and Selection . 99

11.2.1 The Candidate Ranker: prioritizing features and Ranking Candidates – Methodology ... 101

11.2.2 The Mitigation Tool: Identifying and mitigating BIAS from CVs and cover letters – Methodology 104

11.3 Second Block: Trustworthiness – Methodology 106

11.3.1 Activity Overview 106

11.3.2 The ALTAI Requirements 106

11.3.3 The Exercise in practice 108

11.4 Third Block: Learning Needs – Brainstorming 108

12. The International Workshop Co-Creation - Results 110

12.1 First Block: simulating the use of AI based systems in Recruitment and Selection - Findings 110

12.2 Activity One: The Candidate Ranker – Results 110

12.2.1 Exercise execution 110

12.2.2 Strengths and shortcomings of the tool 114

12.2.3 Recommendations to Debiaser developers..... 117

12.3 The Mitigation Tool: Identifying and mitigating BIAS from CVs and cover letters – Findings 119

12.3.1 Exercise execution 119

12.3.2 Strengths and shortcomings of the tool 121

12.4 Second Block: Trustworthiness – Findings 124

12.4.1 Exercise Overview 124

12.4.2 Key concerns over trustworthiness 125

12.4.3 Key solutions for trustworthiness..... 125

12.4.4 ALTAI risks and solutions – AI Experts perspective 126

12.4.5 ALTAI risks and solutions – HR Experts perspective..... 127

12.4.6 ALTAI risks and solutions – Workers and CSO representatives Perspective perspective 128

13. Concluding remarks 130

14. References 132

15. Annex 1 – Facilitation principles & conflict management tips..... 133





16. Annex 2 – Scenarios proposed.....	137
17. Annex 3 – Personas for the first workshop	141
18. Annex 4 – template of cover letter	145
19. Annex 5 – Template for the walking plenary	146
20. Annex 6 – Reporting template for note takers – first workshop.....	147
21. Annex 7 – First workshop report	149
22. Annex 8 - Material for the second round of workshops – first group work.....	152
23. Annex 9 – Template report first group work – second workshop	156
24. Annex 10 – Table for the fourth activity – second workshop.....	159
25. Annex 11 - Slides introducing the Debiaser and the CBR model	160
26. Annex 12 – Second group work report – second workshop	166
27. Annex 13 – Overall report – second workshop.....	169
28. Annex 14 – Simulation tools material	175
29. Annex 15 – The Candidate Ranker: guidelines and instructions	182
30. Annex 16 – The Mitigation Tool: guidelines and instructions.....	188
31. Annex 17 – The ALTAI requirements brainstorming.....	192
32. Annex 18 – Learning needs	195





2. Table of figures

Figure 1 Stakeholders per category in 1 st co-creation workshop	31
Figure 2 Gender of participants in 1 st co-creation workshop	31
Figure 3 Main types of bias identified in first co-creation workshops	33
Figure 4 1st co-creation workshop at SVEN	34
Figure 5 1st co-creation workshop at NTNU	39
Figure 6 1st co-creation workshop at ULEID	42
Figure 7 1st co-creation workshop at FARPLAS	49
Figure 8 1st co-creation workshop at DIGI	53
Figure 9 Positive and negative bias associated to gender and race/ethnicity	62
Figure 10 Words/sentences per category	63
Figure 11 Words/sentences per category and kind of bias	63
Figure 12 Main points concerning fairness in HR recruitment/selection	68
Figure 13 Poster template for the 2nd group work exercise in the 2 nd co-creation workshop	72
Figure 14 Distribution of fields of expertise among participants to the 2nd co-creation workshops	77
Figure 15 Gender distribution in the 2nd co-creation workshops	77
Figure 16 Graph representing the scores of every fairness principle, considering votes and penalties	80
Figure 17 Scores of every fairness principle, considering votes and penalties	80
Figure 18 Interactions between a simulation tool and the user	101
Figure 19 Simplified User Journey with the Candidate Ranker Tool	102
Figure 20 Interface of the Candidate Ranker Tool	104
Figure 21 Simplified User Journey of the Mitigation Tool	105
Figure 22 Interface of the Mitigation Tool	106
Figure 23 Iterations for each group with the Candidate Ranker	111
Figure 24 Must-have Requirements selected in the Candidate Ranker	111
Figure 25 Nice-to-have Requirements selected in the Candidate Ranker	112
Figure 26 Macro-themes of the selected features: distribution	113
Figure 27 Gaussian Distribution of the ranking positions calculated by the Candidate Ranker	114
Figure 28 Detailed distribution of the ranking positions for every group and its iterations	114
Figure 29 Example of the hierarchical decomposition of every feature in sub-categories	117
Figure 30 Iterations of each group with the Mitigation Tool	119
Figure 31 Frequency of the key biases selected by participants in the Mitigation Tool	120
Figure 32 Additional Cross-cutting biases selected by participants in the Mitigation Tool	121
Figure 33 Bias types detected by the Mitigation Tool in the candidates' profiles	121





3. Table of Tables

Table 1 Categories and numbers of stakeholders involved in the two workshops	17
Table 2 Categories and numbers of stakeholders involved in the first workshop	19
Table 3 PROs and CONs of the use of AI systems in recruitment and HR management	21
Table 4 Gender categories for personas in 1 st co-creation workshop	23
Table 5 Structure and description of group work from 1 st co-creation workshop	25
Table 6 Examples of reporting from 1 st co-creation workshop	28
Table 7 Reports and associated language from 1 st co-creation workshop	29
Table 8 Categories and numbers of stakeholders at SVEN's first co-creation workshop	34
Table 9 Key takeaways from the first Italian co-creation workshop	38
Table 10 Categories and numbers of stakeholders at NTNU's first co-creation workshop	39
Table 11 Key takeaways from first Norwegian co-creation workshop	42
Table 12 Categories and numbers of stakeholders at ULEID's first co-creation workshop	43
Table 13 Key takeaways from the first Dutch co-creation workshop	45
Table 14 Categories and numbers of stakeholders at HI's first co-creation workshop	46
Table 15 Key takeaways from the Icelandic workshop	49
Table 16 Categories and numbers of stakeholders at FARPLAS' first co-creation workshop	50
Table 17 Key takeaways from the 1 st Turkish co-creation workshop	53
Table 18 Categories and numbers of stakeholders at DIGI's first co-creation workshop	53
Table 19 Key takeaways from the 1 st Estonian co-creation workshop	57
Table 20 Categories and numbers of stakeholders at BFH's first co-creation workshop	57
Table 21 Key takeaways from the 1 st Swiss co-creation workshop	60
Table 22 Other grounds of discrimination addressed by partners in the spreadsheets	62
Table 23 Categories and ideal numbers of stakeholders involved in the 2nd co-creation workshop	66
Table 24 Structure of the second workshop	66
Table 25 Scenario and persona of the 2 nd co-creation workshop	71
Table 26 Scenario and persona of the second workshop	74
Table 27 Structure and description of group work in the 2 nd co-creation workshop	74
Table 28 The results of ranking the fairness principles	79
Table 29 Emerging needs for the NLP based and CBR Debiaser tools	89
Table 30 External stakeholders at International Co-Creation Workshop	98
Table 31 Project members from each consortium member	99
Table 32 Programme of the International Co-Creation Workshop	99
Table 33 Suggestions for conflict management in co-creation workshops	135





4. Executive Summary

The present document outlines the methodology for organizing two rounds of BIAS co-creation workshops, which took place in the seven countries covered by this action (Italy, Norway, the Netherlands, Iceland, Estonia, Switzerland, and Turkey) and the methodology employed during the international co-creation workshop which took place in Venice. In addition to describing the key methodological approach and the overall decisions made to ensure that the results of co-creation contribute to the design of the Debiaser and Case-Based Reasoning (CBR) systems in WP3 (Chapters [2](#) and [3](#)), this version of the document provides detailed information about the **methods and reporting process** for the first round of workshops (in Chapter [4](#)), the second round (in Chapter [6](#)) and the international workshop (in Chapter [8](#)). Links to templates designed for communication purposes and for group work during the workshops, as well as for reporting in the subsequent phases, are embedded within the text. Facilitation techniques and tips for conflict management are provided in [Annex 1](#).

Additionally, this report presents and analyses the **results** of the first round of workshops (Chapter [5](#)) which engaged 144 active participants, primarily from key stakeholder categories relevant to this phase, including HR officers, workers, and minority representatives/advocates, as well as AI specialists. Also it details the results of the second rounds of workshops (Chapter [7](#)), involving 131 active participants including legal experts and philosopher. Finally, it also incorporates the results of the international workshop (Chapter [9](#)) which involved the active participation of 45 individuals, including stakeholders, partners, and notable field representatives.





5. Introduction and methodological approach

The BIAS project adopts a participatory and co-creative approach to define the requirements for identifying and mitigating bias in AI systems.

In BIAS, the co-creation workshops serve the purpose of providing input for the technological development in WP3 and the exploitation activities in WP6. According to the GA (Grant Agreement), the co-creation process comprises two phases:

1. Providing AI experts involved in WP3 with insights into real-world experiences related to bias detection and mitigation.
2. Shaping an exploitation path for BIAS within WP6, which leverages knowledge and expectations from relevant innovation ecosystems and potential users/buyers of the solutions.

"The value of adopting a multi-stakeholder approach in the design of AI solutions has been well-documented in relevant literature. It is widely recognized that designing these solutions entails not only a multidisciplinary technological effort but also incorporates various other aspects, including social, economic, political, legal, and more. This is due to their potential to have a profound impact on society as a whole ([Leikas et al., 2019](#)). This understanding necessitates the adoption of a multidisciplinary and multi-stakeholder approach, resulting in sustained co-creation between developers and users throughout the technology's development, implementation, and utilization ([Waardenburg, Huysman, 2022](#)).

Three factors are considered crucial for a successful co-design implementation:

1. Stakeholder selection to ensure an appropriate set of participants is chosen.
2. Choice of tools and techniques, where existing co-design methods are adapted to the specific project.
3. Selection of a suitable physical setting to ensure that co-design activities can take place effectively. ([Robertson et al., 2019](#)).

In the BIAS project's co-creation methodology, special attention was dedicated to thoroughly consider the three factors mentioned above, which will be further elaborated upon.

At this stage of the methodology, the following participant profiles were selected in alignment with the multi-stakeholder approach to co-creation: AI specialists, researchers, practitioners, HR specialists, workers, applicants from both academia and the private/industry sector, workers' and minorities' representatives and advocates, as well as philosophers.

During the initial phase of the project, co-creation activities were exclusively planned to serve the purpose of informing WP3 and facilitating the early design of the AI-based Debiaser, specifically in Natural Language Processing (NLP) and Case-Based Reasoning (CBR). Therefore, the initial steps in formulating the co-creation methodology for this phase involved defining its specific goals and identifying expected outputs to provide valuable input for computer scientists in WP3. It was essential to ensure that co-creation did not become a mere engagement and consultation exercise disconnected from the project's overarching goals and the consortium's initial vision.

During the initial phase, SVEN gave careful consideration to defining the methodology's specific goals in collaboration with WP3 task leaders through regular meetings. This required striking a balance between creating engaging hands-on activities for participants and generating valuable input for WP3 within the context of an AI development model. WP3 leaders identified their specific needs as follows:

- Supporting the identification of words (nouns and attributes) and sentences that may lead to bias, particularly in relation to gender and race/ethnicity in selection/recruitment contexts. This involved proposing reformulations to mitigate such biases and generating word lists to be used





in static word embedding for NLP models (further details on the exploratory approach related to the use of the word lists are in paragraph [5.4.2](#)).

- Eliciting knowledge on different interpretations and meanings of the concept of fairness and an equitable selection process, as well as how these concepts are operationalized in various contexts by different stakeholders. The output of this process is intended to inform the CBR models integrated into the Debiaseer and the fairness evaluation of both NLP and CBR-based tools.
- Identifying the ideal features of a Debiaseer Tool for use in the application screening process and early stages of HR recruitment. This involved defining both functional and non-functional requirements to guide the design of the Debiaseer tool.

The specific methodology for the first co-creation workshop was designed to address the needs of the initial expected outcome, which involved generating word lists that might lead to bias and proposing reformulations to mitigate this bias. In contrast, the second workshop and its methodology were centered on assisting in the identification of requirements for the design of the Debiaseer Tool and expanding understanding of fairness in HR concepts. Lastly, the international workshop was specifically structured to collect feedback from stakeholders regarding their perceptions of the Debiaseer; it aimed to conduct a thorough analysis of the system's requirements for achieving the status of a trustworthy AI-driven system.

The methodological guidelines provided to partners, as detailed in the following chapters, encompass crosscutting aspects related to stakeholder engagement. Target groups were identified in alignment with the diverse objectives of each workshop (see Chapters [4](#), [6](#), and [8](#)). All the workshops had their respective agendas, co-creative activities, and techniques.

The use of scenarios as a method to engage participants and foster discussion played a prominent role in both the first and second workshops, and it represented a fundamental element to develop simulations tools in the international workshop. In fact, scenarios are regarded as a valuable tool in the relevant literature ([Leikas et al., 2019](#)) for capturing essential qualitative information from users and stakeholders, which is necessary for the systematic analysis of ethical issues in specific design cases. Scenarios were used in conjunction with 'personas,' fictitious characters representing users with different roles, needs, and diverse attributes. The purpose of working with personas is to start the product development process with the everyday experiences and needs of users in mind ([Nielsen, 2011](#)).

Scenarios and personas were central techniques in the three rounds of co-creation workshops during group activities. In the first workshop, participants were asked to simulate the early recruitment process, which involved evaluating candidates' motivation letters in response to job vacancies announced by HR managers using these scenarios and personas. Subsequently, participants were tasked with reformulating or rewriting excerpts from these motivation letters that had the potential to lead to biased decisions. This approach was designed to generate word lists for WP3 to use in bias detection within static word embeddings (for more details, see Chapter [4](#) on how this was implemented in co-creation activities).

The purpose of this activity was to validate emergent Intersectional Bias Detection (IBD) in Static Word Embedding Association TEST (WEAT) methodologies in different languages, mostly building on previous work from [Caliskan et al. \(2017\)](#) and [Guo and Caliskan \(2021\)](#).

In particular, implicit bias in humans is often measured using the Implicit Association Test ([Greenwald et al., 1998](#)). Such tests can be accessed on the website like of the Harvard project IMPLICIT¹. The tests are available in different languages and cover various topics. In the IAT, human subjects are required to pair two words from different groups, and their biases are measured based on their reaction times. For example, in the IAT on Gender, the test measures whether there is a difference in how male and female terms are associated with math and arts words, among others.

¹ <https://implicit.harvard.edu/implicit/>





Based on the IAT, the Words Embeddings Association Test (WEAT) was developed to measure bias in word embeddings rather than in human subjects ([Caliskan et al., 2017](#)). (Static) word embeddings are dictionaries of human words matched to mathematical vectors in high dimensions, which are used for various computational text analysis tasks. Conclusions about word meanings can be drawn based on the distance between words in the vector space. For instance, the vectors for 'Cat' and 'Dog' will have a closer vector distance than 'Cat' and 'Thunderstorm.'

The WEAT uses the same word lists as the IAT, but instead of measuring reaction times, it uses vector distance (cosine similarity) to determine whether there is a statistically significant bias or not. The null hypothesis is that there is no difference between the two sets of target words concerning their relative similarity to the two sets of attribute words, meaning there is no gender bias within the target word groups.

To conduct bias detection in static word embeddings, we require the same list of words used for an IAT from Psychology. In some cases, these lists may be readily available from our previous research, as well as from the research of others, in certain languages (see below). Initially, these lists can be obtained by translating them (sometimes with adaptations) from other languages. However, our work has indicated that bias can vary across different cultures and languages ([Kurpicz-Briki, 2020](#)) ([Kurpicz-Briki & Leoni, 2021](#)).

In work extending the WEAT ([Caliskan et al. 2017](#), [Guo and Caliskan \(2021\)](#)) developed a method for detecting intersectional bias - attributes associated with members of more than one social group (e.g., African American females, Mexican American males) - in static word embeddings (SWEs). Intersectional Bias Detection (IBD) identifies words that represent biases associated with intersectional groups automatically. This is achieved through a method similar to WEAT. Words whose corresponding vectors are close to those representing an intersectional group, typically characterized by the most common first names within that group, are identified as biases associated with that group. The authors found that the language models they tested exhibited more evidence of intersectional bias than gender or racial bias separately. Therefore, the need to create new lists from the co-creation activities in the first step of the process was identified. These lists are necessary to validate IBD in other languages and to advance with testing based on the state-of-the-art literature in the field, including more recent papers on Emergent Intersectional Bias Detection (EIBD), which refers to biases unique to intersectional groups ([Guo and Caliskan 2021](#)).

Knowledge needs stemming from the other AI model featuring the Debiaser, namely CBR, had a more direct impact on the methodological choices made for the second co-creation phase. This phase was related to exploring stakeholders' opinions and practices regarding fairness in candidates' selection and recruitment processes. The project employs Case-Based Reasoning (CBR) as an alternative to classical Machine Learning (ML) within a Decision Support System to create more transparent decision-making algorithms with a focus on fairness. Designing and developing a CBR-based system involves a different process compared to developing a mainstream ML-based system, and it is seen as a solution to some of the current problems related to fairness.

Existing ML fairness research is limited, which has led to known problems of unfairness in decisions made using classical ML and their approaches to fairness. Fairness is a multidisciplinary concept, and its definition should be sensitive to the context, such as the task, sector, or country. Therefore, defining fairness in AI requires collaboration with non-technical stakeholders, including those who will use the system (e.g., HR professionals), individuals whose lives may be affected by the decisions, government agencies, legal experts, philosophers, and more.





In the context of the recruitment problem, three crucial design tasks are involved in the development process of the Decision Support System (DSS):

1. Deciding how the data will be prepared.
2. Designing and developing the decision-making module.
3. Designing the method for evaluating the system's fairness.

Unfair decisions can occur at each of these stages, and mitigation measures can be applied at each stage. When it comes to the decision-making component, fairness considerations play a significant role. The objective function of an AI system influences the decision-making strategy, and different fairness criteria may require different strategies. Context, including the task domain and country, is a crucial factor in determining fairness requirements. Embedding fairness-related considerations in ML is a complex endeavour ([Mitchell and others, 2021](#); [Saravanakumar, 2021](#)).

For example, if the decision-maker is a hiring company, they may seek to maximize their utility by hiring the best-qualified candidates through a merit-based hiring strategy. In such cases, decision-makers might assume that they are making fair decisions if individuals with the same merit score are treated equally.

However, if the goal includes achieving justice, a different decision-making strategy might be necessary. For instance, specific quotas for females with children could be considered. It's important to note that the notion of fairness is highly context-sensitive and varies across countries, organizations, institutions, and companies.

These fairness constraints must be incorporated into the Decision Support System (DSS) decision-making component and process. This underscores the importance of stakeholder involvement in defining fairness constraints, informed by social, moral, legal, and other dimensions.

In summary, defining and achieving fairness in AI is a complex endeavor, particularly in the context of recruitment. It emphasizes the need for interdisciplinary collaboration and context sensitivity in addressing fairness issues throughout the AI development process. As a result, the methodology elaborated for the second round of co-creation workshops adopted scenarios and personas as techniques to elicit knowledge and definitions of fairness to be used in CBR model design and the fairness evaluation of the overall system.

Applicable to both ML models used within the project, the second round of co-creation workshops centered on eliciting multi-stakeholders' expectations, perspectives, and reflections regarding the potential requirements of a Debiaser. To facilitate this, activities inspired by the 'future-state journey map' technique² were integrated in the relevant methodological guidelines. The guidelines for the second round of workshops were designed to specifically aid the process of identifying desirable requirements for the development of a Debiaser tool and a CBR system. The technical work in WP3 will incorporate the findings from these workshops.

The requirements were identified by simulating a 'recruiter journey' in the process of selecting a candidate for a specific job vacancy. Although this technique is typically used by companies to enhance their customers' experience, it was adapted to meet the specific context and requirements of the second round of co-creation workshops (for more details on the methodology of these workshops, please refer to [Chapter 6](#)).

Building upon the activities conducted in the two rounds of workshops, the international workshop featured a comprehensive exercise, aiming to gather feedback on simulated tools replicating future CBR and word-embedding based systems, aligning with the ALTAI paradigm on AI system trustworthiness, and

² <https://www.mindtools.com/aijjpy/designing-future-state-customer-journeys>





addressing effective training package development (for more details on the methodology the co-creation workshop, please refer to [Chapter 8](#)).

The designed simulations replicated all stages of a recruitment process, requiring participants to consider various text data used in the selection phase, which include candidates' profiles, cover letters, resumes, job offers, and the company's policies and values. Using a simulation of a system in a co-creation workshop can be a highly effective strategy for several reasons. Firstly, it is proven to enhance the understanding of the concepts explained: simulations provide a visual and interactive representation of a system. Participants can gain a better understanding of the complex relationships, processes, and dynamics within the system. This hands-on experience can enhance comprehension compared to theoretical discussions alone ([Ruhl, Richter et al., 2014](#)). Secondly, it promotes experiential learning: simulations create a scenario-based environment, where users can safely experiment with different situations and observe the consequences of their actions, also allowing the developers to reach their own goals and to observe direct reactions from the interaction with the tools. On balance, results indicate that games and/or simulations have a positive impact on both users and designers ([Vlachopoulos and Makri, 2017](#)). Additionally, simulations are often used in teaching contexts to facilitate collaboration: participants and students can explore different scenarios, assess potential risks, and identify opportunities without real-world consequences. This encourages creativity and experimentation ([Ruhl, Richter et al., 2014](#)), ([Lee et al., 2018](#)). The idea of developing an online flexible simulation system came from the fact that in the previous workshops it was noted how differently participants would react under different perspectives: since simulations allow for experimentation, participants can iterate and refine their approaches based on the outcomes they observe. This iterative process mirrors real-world problem-solving and improvement cycles, fostering a culture of continuous learning and adaptation. The main challenge for participants in the second round of workshops was grasping the concepts of the NLP and CBR tools of the Debiaser, so it was one of the main priorities to help in visualizing intricate relationships and dependencies within the systems, making it easier for participants to grasp the holistic view of the Debiaser. Finally, the international co-creation workshop was designed with the purpose of extracting decision-making choices for the proof of concept of the Debiaser. The outcome of the workshop was positive in this sense: the simulation exercises and their consequent feedback-collection activities, assisted participants in influencing development decisions by allowing them to see the potential outcomes of different choices in a realistic setting; participants had the opportunity to reflect on their interactions with the systems, and were able to anticipate many challenges and develop strategies for a fair and trustworthy implementation of the Debiaser.





6. Cross-cutting methodological aspects for the two workshops on engagement and preparation

6.1 Targeted participants/stakeholders

As previously mentioned, the first co-creation phase is directly linked to the early stages of technology development, and as such, the workshop structure was designed by SVEN in collaboration with WP3 task leaders.

To achieve the aforementioned results, the team agreed that the workshops should incorporate specific group work, each prioritizing specific stakeholder categories. While the Grant Agreement in section T2.5 generally mentions that the workshops will mainly involve AI specialists, researchers, students, practitioners, HR specialists, workers, applicants from academia and the private/industry sector, workers' and minorities' representatives and advocates, with the same participants in both workshops, a change in the initial plan was proposed due to the distinct objectives and expected results for each workshop during the project's implementation.

For the first workshop, the ideal participants were drawn from the following stakeholder categories, listed in order of priority and numerical participation:

- ➔ Workers and workers' representatives (e.g., trade unions).
- ➔ Representatives of civil society organizations (e.g., associations, NGOs), networks, organizations advocating for equality and inclusion, and combating discrimination, particularly related to gender and race.
- ➔ HR officers and networks, associations of HR specialists.
- ➔ AI specialists.

In contrast, the second workshop primarily involved potential users of the Debiasser system (HR officers) and AI/tech experts who could provide input on its design, with some participation from minority and workers' representatives. Other stakeholders, such as philosophers and human rights and legal experts in human rights and labour law, were also engaged, albeit to a lesser extent.

To ensure valuable content results, the number of participants for each workshop was limited to 24 attendees, with a goal of involving a minimum of 35 different stakeholders across both workshops to meet the set Key Performance Indicator (KPI).

The first round of co-creation workshops took place between June and September 2023. Each partner conducted two workshops, one between M8-M9 (June – July 2023) and the other between M10-M11 (August – September 2023). Despite having different purposes, specific content, and activities, the two workshops shared a similar structure and agenda, including an initial introduction, a first discussion activity, group work, and a final networking moment, which took the form of a networking aperitif, lunch, or dinner depending on the time the workshops were organized. Each workshop lasted for approximately 4 hours.

As already mentioned, both workshops involved around 24 people each.

Each partner had 4900 euro available in their BIAS budget under “Other Direct Costs” to use for the organization of the workshops. The amount was used in flexible ways. Beside the costs incurred for the catering, the room rental, materials, etc. partners could use the resources for incentivizing the participation in the workshop of particularly relevant stakeholders and motivating them in assuming a more engaged role by inviting other organisations/people to join, by contributing to the concrete organisation of the workshop. Further details are provided in the following paragraph.





6.2 Preparatory and engagement activities

Preparatory activities mainly involved the identification and setup of related arrangements, including:

1. Facilitators and rapporteurs to run the workshops.
2. Stakeholders to invite to the workshops.
3. Securing an appropriate location.

Regarding the first point, it was recommended that each partner identified two individuals to act as facilitators during the workshops. Facilitators did not need to be experts in AI, HR, employment, or discrimination, but they should preferably have experience in at least some of the following activities:

- Applying participatory methods to facilitate group discussions.
- Providing hands-on and participatory training or capacity building.

Facilitators were expected to possess strong communication and active listening skills, as well as an understanding and familiarity with issues related to gender equality, diversity, and intersectionality. Facilitators could have been either internal members of the partners' teams/organizations (from the same or different departments) or external professionals who were contracted for this purpose. [Annex 1](#) of this document provides concise guidelines for facilitators, including facilitation principles and tips for conflict management.

It is worth to underline that, according to the GA, some partners (BHF, DIGI and FARPLAS) had dedicated resources (7.000 euro) for hiring facilitators.

Smart Venice organized two online training sessions for partners to prepare facilitators and provide them with the necessary knowledge and tools to organize and conduct the workshops. The first training session, in preparation for the first co-creation workshop, was scheduled for May 25, 2023, while the second session took place on July 19, 2023. At least one of the facilitators identified by each partner attended these sessions.

Facilitators also assumed the role of rapporteurs during the group work activities. Since four group work sessions were conducted during the workshops, an additional two individuals to serve as rapporteurs needed to be identified. Rapporteurs were not required to possess specific skills or competencies. They could also be students, but in any case, they needed to be familiarized with the project, its objectives, and the reporting process, which they were tasked to use.

Regarding the engagement of relevant stakeholders, the following steps were suggested to partners:

1. Thoroughly identify potential stakeholders belonging to the various categories to be involved, as listed in Table 1 below:

Table 1 Categories and numbers of stakeholders involved in the two workshops

First workshop	Second workshop ³
8 Workers and workers' representatives (e.g., trade unions)	10-12 HR officers and networks, associations of HR specialists
8 Representatives of civil society organisations (e.g., associations, NGOs), networks, organisations advocating for equality and inclusion and fighting against discriminations (in particular, but not exclusively related to gender and race)	4 AI specialist, practitioners, academics, researchers, students

³ Numbers of ideal participants per category will be provided in the second version of the present methodology.





4 HR officers and networks, associations of HR specialists, preferably already active on gender/diversity & inclusion issues	2-4 Workers and workers' representatives
4 AI specialists	2-4 Representatives of civil society organisations (e.g., associations, NGOs), networks, organisations advocating for equality and inclusion and fighting against discriminations
	2 philosophers
	2 legal experts in human rights and/or labour law

It was recommended to identify a minimum of 15-20 stakeholders per category to meet the expected number of participants for each type as indicated in the table above.

It was also expected that the invited stakeholders would participate in the National Labs and also display a high level of interest and motivation in engaging with the workshop's topic. SVEN emphasized the importance of having as diverse a group of participants as possible.

In order to elaborate their invited stakeholders' list, partners relied on:

- the stakeholders' mapping conducted within WP7
- participants in the interviews conducted within T2.3
- existing contacts/networks they already had in place in the frame of other projects/collaborations.

Two lists of stakeholders were suggested to be produced: a "plan A" with the ideal composition of stakeholders to be invited and a "plan B" with other pre-identified potential stakeholders to be contacted if needed.

2. Send dedicated invitation emails to previously identified stakeholders. Partners could adapt and use the text available at the following link to engage stakeholders in participating in the first workshop.
3. In case the 24 participants were not reached through dedicated email invitations, partners proceeded with social media announcements using templates available in Teams.

Partners were advised to consider incentivizing stakeholder participation by offering a fee, utilizing a portion of the budget allocated for organizing the workshops (€4,900). Partners could decide to provide a fee to all stakeholders or select individuals based on previously identified conditions. These conditions might include stakeholders:

- Whose participation was deemed particularly relevant.
- Traveling from other cities, incurring transportation and accommodation expenses.
- Whose participation would require taking an unpaid day off from work.

Furthermore, partners identified suitable **locations** that could accommodate 24 people and offer the flexibility to work in smaller groups. It was suggested that during the "group work" sessions, four smaller groups be formed, each consisting of approximately six people. Additionally, venues needed to ensure accessibility for individuals with reduced mobility requirements. On accessibility matters in general, it was recommended to inquire about any special needs related to visual or auditory impairments during the registration process to address them appropriately during the meeting.





7. The first workshop's specific co-creation methodology

7.1 Workshop's agenda and target

As already mentioned in the previous chapter the workshop lasted approximately 4 hours and was structured as follows:

- Introduction & BIAS presentation (15 min)
- Bias in HR and recruitment: open discussion (45 min)
- Group works (2 hours and 15 min - with a coffee break in between)
- Coffee/happy hour/lunch & networking (approximately 1 hour)

The primary goal of the first workshop was to identify the categories of wordlists required for WP3 in relation to bias detection in static word embeddings. This objective was accomplished by initiating discussions with and among participants about the types of biases likely to manifest in the recruitment process for various job roles. Two dimensions were analyzed concerning discrimination grounds and axes of inequalities: gender and ethnic/cultural background. However, other potential biases were also explored.

As previously mentioned, the first workshop involved the following stakeholder categories:

Table 2 Categories and numbers of stakeholders involved in the first workshop

Type	Ideal number
Workers and workers' representatives (e.g. trade unions)	8 people
Representatives of civil society organisations (e.g. associations, NGOs), networks, organisations fighting against discriminations (in particular, but not exclusively related to gender and race)	8 people
HR officers and networks, associations of HR specialists preferably already active on gender/diversity & inclusion issues	4 people
AI specialists	4 people

When splitting into smaller breakout sessions for group work, it was advised to maintain the same proportions to ensure balance.

The workshop aimed to engage a minimum of 24 individuals from the aforementioned categories. Workshops were conducted in person, as the networking aspect was recognized as one of the primary incentives for participants to attend.





7.2 Introduction & BIAS presentation

The first 20 minutes of the workshop were dedicated to introducing the BIAS project. Each partner was required to prepare a few introductory slides in the language adopted for the workshop. Additionally, information related to WP3's work on the Debiaser was presented using slides prepared by.

During the introductory section, participants were provided with information about the project, its objectives, the role of co-creation, and how the workshop's results would be utilized. Specifically, the following information was shared.

BIAS next steps:

- A **second co-creation workshop** scheduled to take place **between August and September 2023**, focusing on the desirable features of a **Debiaser tool** (partners could disclose the date if already scheduled).
- An **international workshop** to be organized in **December 2023** in Venice, bringing together project partners and stakeholders to discuss the results of the two workshops.
- A final public deliverable analysing and reporting the results of the workshops, which would inform the work of the technical partners in the project.
- Participants who were not yet aware of this opportunity were encouraged to subscribe to the BIAS Labs to stay updated and receive invitations to upcoming activities.

7.3 Panel discussion

It was recommended that each partner identify two to four individuals among the stakeholders participating in the workshop who would be available to take part in a panel discussion on the topic of "bias in HR and recruitment and definitions and meanings of fairness in decision making related to recruitment." The panel was suggested to be structured around a set of pre-prepared questions or discussion points, to be moderated by a facilitator. Ideally, the panel would include a representative from each stakeholder group present at the workshop, or at least representatives from the two prioritized groups: workers and worker representatives, and representatives of civil society organizations.

Below is the guideline provided to facilitators who chaired or moderated the workshops, based on the Italian case. It was presented as a possible blueprint to be adapted to each context, aiming to reference and integrate current topics in the national public and expert debate on the relevant subjects.

Introduction: Italian companies still have a long way to go in fully embracing diversity, equity, and inclusion policies. A recent survey titled "Equality, Diversity, and Inclusion Research Italy," conducted by Workday, a leading company in corporate cloud applications for finance and human resources, in collaboration with Sapio Research, explored this issue with the participation of 301 HR professionals and Italian business leaders from both multinationals and SMEs. The findings from the report reveal some alarming aspects, but they also offer room for optimism.

According to the research, one in three companies in Italy, equivalent to 36%, either denies or downplays issues related to equity, diversity, and inclusion, particularly concerning the acceptance of differences in gender, ethnicity, religion, sexual orientation, age, and social background. On the other hand, 35% of the respondents indicated that their organizations adopt commendable practices for managing diversity, while 25% stated that their company encourages dialogue and mutual acceptance among employees. Furthermore, 75% of the companies have allocated a budget for Equality, Diversity, and Inclusion (EDI) policies.

Question 1: Based on your knowledge and experience, do we have good reasons to be optimistic or should we rather be concerned with the situation in our country?





The use of AI in Human Resources management and in recruitment processes is becoming widespread as a source of innovation that could also support EDI policies, while it is controversial in many respects: take the use of Chat GPT, it has several PROS and CONS.

Table 3 PROs and CONs of the use of AI systems in recruitment and HR management

PROS	CONS
Get rid of repetitive HR tasks like scheduling interviews, drafting bulk emails.	Limitations in complex situations: it fails assisting HR executives in situations where human judgement and empathy are required.
Speed up the process by providing responses and filling up the appraisal forms with candidate details for the management to take quick decisions. Help HR executives in improving on their communication strategies.	Compromised privacy: there are pretty much chances that AI applications can go rogue and saving sensitive employee information becomes next to impossible.
Enhance data-management with real time updates and insights.	Addiction/Heavy Dependence on Technology, lost of the human touch.

Question 2: What is your opinion and/or experience regarding the use of AI systems in recruitment and Human resources management in general? Are the PROs too enthusiastic on tech-innovation? What points of attention would you advise to balance the CONSs?

Linking back the topic of AI use in recruitment and HR management and Equality/Diversity Policies, scholars and activists are warning precisely that AI risks reproducing and strengthening bias and inequalities. Well known is the case of **Amazon's** recruiting software which ended up discriminating against women ([Dastin, 2018](#)). It was found that biased recruitment data of the company of the previous ten years were used for training the software, which replicated human mistakes. Similarly, **LinkedIn** discovered that its "recommendation" algorithms, which were used to match candidates with job opportunities, produced "distorted" results, favouring male candidates over women ([Wall, Shellmann, 2021](#)).

Question 3: What is your view on the role that AI based technology can play to favour or to hamper EDI in hiring processes in particular?

In BIAS, we aim at designing a "fair" and trustworthy AI system able to detect and mitigate bias in recruitment, but what is a fair hiring process and procedure in your view?

Question 4: How would you define it and to what extent such definition is context dependent in your view?

Finally, a recent study ([Nursky, Hoffmann, 2022](#)) has shown that "Meaningful workers participation in the adoption of workplace AI is critical to mitigate the potentially negative effects of AI adoption on workers, and can help achieve fair and transparent AI systems with human oversight. Policymakers should strengthen the role of social partners in the adoption of AI technology to protect workers' bargaining power".

Question 5: What is your view on this? How participation of workers and social partners but also civil society organizations representing minorities can contribute to influence and oversee the use of AI in recruitment and make it fairer?





Considering the overall allocated time for the panel discussion (45 minutes) and depending on how many panellists were identified, it was suggested to predefine how much time each one had to discuss on the proposed topic. It was suggested that tentatively, five to ten minutes should be allocated for each speaker, making sure to leave at least 20 minutes for an open discussion and interaction on the topics addressed with the overall audience.

7.4 Co-creation group work

The workshop featured a core co-creation activity in the form of group work, as mentioned in the [introductory chapter](#). This group work incorporated scenarios and "personas" as methods to engage participants and stimulate discussions, aiming to identify relevant wordlists and wordlist categories. Specifically, the group work sought to uncover words associated with intersected categories within a recruitment process, particularly when considering gender and race as potential grounds for discrimination. It is worth emphasizing that a **comprehensive definition of "bias" was adopted⁴**, encompassing **both positive and negative biases within the exercise**.

The participants were divided into four groups, each consisting of six individuals, and were provided with a scenario and descriptions of different personas. The composition of each group was diverse, ensuring that at least one HR officer/specialist and one AI specialist were included in each group.

Scenarios & "personas"

All four groups worked with the same scenario, which was chosen from a set of four different scenarios featuring fictitious job offers. These scenarios were provided as options for partners to select during the workshop. However, if the pre-designed scenarios did not align with the specific context or job market of the partners, they had the flexibility to create a new scenario that better suited their needs.

The four proposed scenarios included Job Offers descriptions from different companies/organizations in a variety of sectors: an iron/steel industrial company, a research institute, a tech company, a private school.

The scenarios are available in [Annex 2](#).

It's important to note that the recruitment processes described in the scenarios do not incorporate the use of AI technology. Partners were advised to ensure that participants clearly understood this aspect.

It is important to note that the selected sectors were chosen to address both "horizontal" and "vertical" segregation phenomena. "horizontal segregation" refers to the concentration of one gender in specific fields of education and occupation⁵, while "vertical segregation" pertains to the concentration of women and man in distinct grades, levels of responsibility or positions, as defined by EIGE⁶. A main distinction was made between STEM professions and care related and EHW (Education, Health and Welfare) professions in which gender/race and ethnicity gaps are well documented. In particular, STEM professions are typically male-dominated fields as shown by the recent She Figures 2021 Report⁷ and relevant literature ([Giancola, De Vita, 2017](#)), while care related and educational professions (with educational professions excluding the ones related to HE) are typically female dominated, and very often racialized as well ([Equinet, 2022; Marchetti, 2022](#)). STEM professions encompass both 'science' and 'mathematics,' as indicated by the relevant literature on Word Embedding Association Tests (WEAT) by [Caliskan et al., \(2017\)](#). These

⁴ See definition given by Merriam-Webster available at this link <https://www.merriam-webster.com/dictionary/bias#dictionary-entry-3>

⁵ <https://data.consilium.europa.eu/doc/document/ST-12709-2017-ADD-2/en/pdf>

⁶ https://eige.europa.eu/publications-resources/thesaurus/terms/1243?language_content_entity=en

⁷ <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/67d5a207-4da1-11ec-91ac-01aa75ed71a1>





categories served as the foundation for defining the human-created wordlists, which were the intended output of the first workshop as discussed in [Chapter 2](#).

In instances where partners identified other sectors more relevant to their specific context, they were encouraged to create 'new' scenarios. In doing so, it was recommended that these scenarios align with employment sectors characterized by both gender and race vertical and/or horizontal segregation, as previously discussed.

Furthermore, partners were advised to choose existing job offers, ensuring they included essential elements such as the position, type of organization offering the job, contract details, expected requirements and qualifications, responsibilities, and tasks.

In the second step, the focus was on identifying four key "personas," which are fictitious characters designed to emphasize two intersecting dimensions, with priority given to gender and race/ethnicity. To maintain consistency across all four groups in the same workshop, the personas were designed with the needs of WP3 in mind. These personas were created by using different combinations of selected aspects for gender and ethnic group. For each selected gender and race/ethnicity category, both 'positive' and 'negative' connotations were introduced. This approach simplified the personas but ensured that the generated wordlists could be used effectively within existing bias-word embedding tests (i.e. WEAT, as per [Caliskan et al. 2017](#)).

While the preference was for all partners to focus on gender and race/ethnicity as intersecting dimensions, it was also an option for partners to consider other dimensions if they were more relevant and representative of minority communities facing discrimination in their respective countries. These other dimensions could include religion, disabilities, sexual orientation, among others. The goal was to include both majoritarian and minoritarian features. For instance, if a partner chose to focus on the dimension of disability, the minoritarian profiles could have features like "differently-abled," "physically challenged," or "mentally challenged," while the majoritarian profiles would be labelled "without disabilities/impairments."

Regarding the gender dimension, SVEN suggested that partners consider a non-binary definition of gender to enhance the intersectional research approach. Various gender categories were presented as options, but if the cis/non-cis distinction was deemed not relevant or potentially challenging in a specific country, the binary option was still provided.

Table 4 Gender categories for personas in 1st co-creation workshop

Approaches to gender	Persona 1	Persona 2	Notes
Binary approach	Man	Woman	
Non-binary ⁸ approach	CIS person	Non-CIS person	If person 1 is CIS man, persona 2 will be non-CIS man; if persona 1 is cis-woman, persona 2 will be non-CIS woman

When referring to "non-cis persons", the intention was to include individuals who identify as queer, non-binary, or trans⁹.

⁸ In order to include a non binary definition of gender, we refer to the term "Cisgender" identifying a person whose gender identity corresponds with the sex the person had or was identified as having at birth, independently from their sexual orientation. On the contrary, the term "non-cisgender" refers to a person whose gender identity do not correspond with the sex the person had or was identified as having at birth <https://www.merriam-webster.com/dictionary/cisgender>

⁹ In case of trans persons it is not relevant whether they identify themselves as man or woman, this could be just highlighted in the persona's profile.





As far as the race/ethnicity dimension is concerned, the following categories were included, aligned with those identified and used in the BIAS survey (T2.3):

- Black,
- Latin American,
- Asian
- Middle eastern
- North African
- Roma
- White

Partners were asked to choose two race/ethnicity categories only to work on within the workshop: white (as the “majoritarian”/“mainstream” race/ethnicity category) and another category which was considered relevant in each country. A choice on this could be made based on different criteria such as:

- national data on migration and/or race and ethnicity discrimination in each country;
- presence of stakeholders in the group that represent a particular minority group.

After choosing the race/ethnicity categories, partners needed to elaborate 4 profiles (one per each working group) intersecting race/ethnicity categories with the gender-related ones.

For instance, in case the relevant race/ethnicity category in a given country was identified as being “black”, 4 profiles could be chosen out of the following:

- 1) Black cisgender man
- 2) Black cisgender woman
- 3) White cisgender woman
- 4) White cisgender man
- 5) Black non cisgender (queer, non-binary or trans) person
- 6) White non cisgender (queer, non-binary or trans) person

Partners had to make sure that black and white were included, while matching gender profiles. For example, four groups could be identified as follow (with adopting a binary definition of gender):

- Black cisgender man
- Black cisgender woman
- White cisgender man
- White cisgender woman

If partners decided to include a non-cisgender component within the four personas, they needed to select the most relevant cisgender profile based on the chosen scenario. For example, assuming partners selected the tech related scenario, then they might have wanted to choose a “cisgender man” profile, since it represented the majoritarian group in tech if they believe that working with positive bias would be more stimulating in their context. Therefore, in such case, the following selection could be made:

- Black cisgender man
- Black non-cisgender (queer, non-binary or trans) person -> it trans, identifying as man
- White cisgender man
- White non-cisgender (queer, non-binary or trans) person -> if trans, identifying as man

On the contrary, assuming partners selected the education related scenario, then they might have wanted to choose a “cisgender woman” profile, since it represents the majoritarian group in education. Therefore, the following selection could be made:

- Black cisgender woman
- Black non-cisgender (queer, non-binary or trans) person -> it trans, identifying as woman





- White cisgender woman
- White non-cisgender (queer, non-binary or trans) person -> if trans, identifying as woman

The examples of “personas” profiles that were made available to partners can be found in [Annex 3](#). When elaborating the profiles, partners needed to fill in the sections “previous work experiences (including career progression)”, “education” and “hobbies/sports and personal attitudes” (including work ethics), other skills and languages according to the “scenario”/job offer that partners decided to work upon, so to ensure there was a fit between personas and vacancy (albeit not necessarily a full-fit).

Group work development

Facilitators split the participants into 4 groups, 6 people per group: ideally 1 HR officer, 1 AI specialist, 4 workers/workers representatives/representatives of civil society organizations, NGOs, networks per group.

4 rapporteurs had to be identified by partners and be in charge of notes-taking. The 4 rapporteurs could have been the 2 facilitators plus two additional people of the partner organization (e.g. students), or else 4 rapporteurs could undertake that role exclusively.

The role of the note takers was crucial for the successful reporting of the workshop’s results. Indeed, they needed to be as detailed as possible when taking notes during the group work in order to permit facilitators/researchers to extract the relevant words/sentences from the final report (see paragraph [4.5](#)).

The facilitators shared with each group the material made available by Smart Venice, containing:

- The scenario/job offer selected ([Annex 2](#))
- 4 “personas” profiles, completed as per described above ([Annex 3](#))
- 4 “personas” profiles, completed as per described above, but with no picture
- a template for the cover letter ([Annex 4](#))
- a template for the walking plenary session ([Annex 5](#))
- a reporting template (for the note taker) ([Annex 6](#))

All the listed templates had to be downloaded, translated in local language, and printed (except for the reporting template for the note taker which can be downloaded and filled using a laptop). The template for the walking plenary session had to be printed in a poster format (A1).

It was also recommended to translate and print the structure of the exercise below. Sticky notes should have also been provided to the working groups. The group work lasted around 2 hours and was structured as follows:

Table 5 Structure and description of group work from 1st co-creation workshop

<p>First activity - discussion on the job offer (15 min)</p>	<p>The HR officer’s role in this part of the workshop is to go through the scenario/job offer and provide insights regarding:</p> <ul style="list-style-type: none"> - which are the prerequisites/expected skills and competences - which is the ideal profile according to the offer text - which elements she/he would expect to find in a successful cover letter. <p>All the other group’s members listen and a short discussion with all participants follows on potential bias deriving from the job offer’s formulation. The rapporteurs report on the results of the discussion (see details in the reporting process below, paragraph 4.5).</p>
<p>Second activity - elaboration of the cover letter (30 min)</p>	<p>The HR officer receives the persona’s profile without the picture of the candidate. The other participants receive the persona’s profile with the picture. Participants go through the candidate’s profile received and briefly question on which kind of biases a person having that profile could face.</p>





	<p>Then, all participants, except the HR officer, work collaboratively to elaborate a cover letter based on the job offer and the profile of a fictitious candidate provided (“persona”) and elaborated in advance by facilitators.</p> <p>Participants should not focus on bias when preparing the cover letter, on the contrary, they should elaborate it trying to “put themselves into the persona’s shoes”.</p> <p>The clarity and legibility of the cover letter are crucial for subsequent reporting. Therefore, the group should select a participant with legible handwriting to write it. The rapporteurs fill in the report template highlighting the main discussion points, the topics that received the most attention and any differing viewpoints that arose during the discussion.</p>
<p>15 minutes coffee break</p>	
<p>Third activity - discussion on the cover letter (20 min)</p>	<p>The group reads the cover letter and discusses with the HR officer on the following questions using sticky notes:</p> <ul style="list-style-type: none"> - Are there any risks of bias coming from the cover letter? Which kind of bias (e.g. gender, race, age, disability, etc.)? - Are there any specific bias coming from the picture of the candidate? - Which are the words/sentences that could lead to bias? Which are the associations to those words/sentences that make them lead to bias? (for instance, the word “children” is not biased per se, but if associated to “remote working” it could lead to bias) - How can these risks affect the decision of a recruiter? - Are the biases related to the work or family/private life sphere or any other category? <p>The group receives clear indication to focus on both positive and negative biases. The rapporteurs report in detail on the results of the discussion highlighting if bias are positive or negative.</p>
<p>Fourth activity - cover letter rephrasing (20 min)</p>	<p>After the discussion, the entire group collaboratively revises the cover letter to rephrase sections that may contain risks of bias. All the outputs of the group works are attached to a pre-designed board/poster.</p> <p>The rapporteurs fill in the provided template, highlighting the main discussion points, specifying the topics that received the most attention, and noting any differing viewpoints that emerged.</p>
<p>Walking plenary (20 min)</p>	<p>Each group presents its result to the other groups using the posters and post-its, explaining:</p> <ul style="list-style-type: none"> - The job offer. - The fictional character profile. - The words/sentences of the cover letter that were at risk of bias and how they were rephrased. - Any other relevant outcome of the work.

Time management during the group work was crucial to ensure the completion of all activities. Partners had various options to manage time effectively:

1. Centralized Time Management: Someone from the hosting partner's staff guided the groups by giving signals about different time slots on a slide, accompanied by a gentle sound when each time slot had expired.
2. Rapporteurs as Time Managers: The rapporteurs could take on the role of managing time, ensuring that the group followed the schedule.





3. Assignment within the Group: The group itself could designate one of its members to manage time and keep the activities on track.

To summarize the results of the group work, partner teams prepared a report. Rapporteurs took detailed notes during the activities using a specific template in the national language available in [Annex 6](#). Additionally, a comprehensive final report in English was drafted immediately after the workshop (details on the reporting process are provided in a dedicated paragraph below). These reports were refined and finalized promptly to minimize the risk of misinterpretation, especially since no recordings were made. Wordlists were later extracted from these reports to identify relevant terms and phrases.

7.5 Reporting process

Partners had to report on the overall results of the workshops and especially of the group works using the templates made available by Smart Venice:

- a document template for reporting the results of the workshop as a whole ([Annex 7](#)), with the translated templates of the four group-work reports as annexes;
- a spreadsheet/matrix for reporting wordlists.

The document was designed to present the results of both the panel discussion and the group work, using information from the rapporteurs' reports (as explained in the paragraph above). This document did not include direct quotes or information that could identify the participants, ensuring full pseudonymization of personal data.

The accompanying spreadsheet followed a specific format. It was structured as a matrix that crossed intersectional dimensions with other selected relevant categories. In this matrix:

- Intersectional categories were in column B. These categories corresponded to the personas' profiles identified by partners and assigned to the groups. Two primary intersectional dimensions were considered: gender and race/ethnicity.
- The other axis of the matrix covered various word categories. These categories guided the selection of wordlists in combination with intersectional identities.

Words' categories were determined based on a combination of existing literature on bias in word embeddings and studies and literature on gender and diversity inequalities and discrimination. These words categories were further organized into sub-categories to facilitate the analysis:

- Career and family issues -> this category includes words and attributes related to both career and family aspects. It was derived from relevant literature on WEAT ([Caliskan et al., 2017](#)) and its crucial for identifying gender and race structural inequalities ([Wharton, 2012](#)). Under the "career" sub-category, terms related to career progression path, career-related skills, and education would be included. The "family issues" sub-category encompasses words and attributes associated with family members, sentimental life, domestic and care work, and work life balance.
- Work ethics -> This category pertains to words and sentences related to behavioral rules and values that contribute to creating a positive work environment and achieving high-quality results. Identifying bias within this category is important for understanding how bias may affect workplace conduct and ethical standards¹⁰.
- Personal information: this category encompasses two sub-categories:
 - Personal attitudes and other skills & knowledges: Words and phrases that reflect individuals' personal beliefs, values, and attitudes, as well as terms related to skills,

¹⁰ <https://harappa.education/harappa-diaries/work-ethic-meaning-definition-and-importance/>
<https://www.personio.com/hr-lexicon/work-ethic/>





- knowledge, and competencies beyond career-specific qualifications. The knowledge of languages was also considered and falls within this sub-category.
- Hobbies & leisure -> Words and phrases associated with individuals' interests, pastimes, and leisure activities. This category is especially pertinent given the influence of gender and race-based patterns on leisure activities and interests ([Henderson, 2013](#); [McDonald & Shelby, 2017](#)).

In addition, an open “other categories identified by the groups” was included to allow for other intersectional axis of inequalities and related bias to emerge, so to balance the indication to stick to two main discrimination grounds (mainly gender/race and ethnicity) as requested by the emerging algorithmic modelling needs from WP3.

Based on the Workshop report, the matrix/spreadsheet was filled in.

Words were meant to include both substantives and adjectives. Full short sentences could also be included.

In the first sheet (named “overall”), partners needed to specify which of the identified words and sentences from the group works, related to the various subcategories listed above (row 3 of the grid), were associated with the four intersectional categories (column B of the grid) for the chosen fictitious characters.

These words and sentences had to be provided in the corresponding cells, both in local language and in English (two cells for each intersection, the upper for words/sentences in local language and the other for their translation in English), separated by a “;” (for instance, “good; migrant origins”).

Partners were also required to indicate **associations between the identified words/sentences or between one word/sentence and another that could lead to bias**. For instance, “children” could lead to negative bias if associated with “remote working,” or “yoga” could be biased if associated with “man.” It was emphasized that words/sentences should be listed **in the same order in both the local language and English**. Additionally, different colors were to be used to denote “positive” or “negative” bias (**green for “negative” and blue for “positive” bias**).

See the examples below:

Table 6 Examples of reporting from 1st co-creation workshop

	Career	Family	Personal info	
			Personal attitudes	Hobbies/leisure
Black man		bambini (lavoro da remoto)		yoga (gestione dello stress)
		children (remote working)		yoga (stress management)

The spreadsheet/matrix included sub-sheets for each different individual profile chosen. In these sub-sheets, the word/attributes/sentences within the various dimensions were further categorized as: 1) potentially bias-generating 2) controversial (if no agreement is found in the group).

The re-formulated versions, attempting to avoid or mitigate bias, were also included. For potentially bias-generating and controversial words, partners were encouraged to report full sentences if they better facilitated the identification of how the combination of words could potentially lead to biased interpretations. It was important to explain the rationale behind adding comments in the spreadsheet.

As rapporteurs and those responsible for writing the workshop reports typically did not overlap, it was strongly advised that rapporteurs made themselves available to answer any questions or address interpretative doubts that might arise during the report drafting process.





The two reports had to be downloaded, filled in and sent to Smart Venice together with the reports of the four group works by **the end of July 2023**.

As far as the languages of the reporting template were concerned, the following table recaps in which language they had to be produced.

Table 7 Reports and associated language from 1st co-creation workshop

Type of reporting template	Language
Group works reports (4 reports)	Local language and English
Cover letter	Local language (to be translated in English only if needed/requested at a later stage by SVEN/BHF)
Overall report for reporting the results of the workshop as a whole	English
Table for wordlists	To be filled with words/sentences both in local language and English





8. Implementation of the first co-creation workshops: results

8.1 Overview of the conducted workshops

In June and July 2023, seven partners (SVEN, NTNU, ULEI, HI, DIGIO, BFH, and FARPLAS) conducted the first round of co-creation workshops. The specific dates for each workshop are reported in their respective paragraphs (see section [5.3](#)). In total, 178 people registered, while 144 actively participated.

Partners made significant efforts to engage workshop participants. Engagement began with partners' personal networks and contacts and continued through email invitations sent to various local and national stakeholders across different categories. Each partner sent hundreds of emails to promote the initiative, and news and posts were also published on the project's and partners' organization websites and social media platforms. These efforts were coordinated and supported by Work Package 7 on dissemination and communication.

The strengths of the workshops, as outlined below, were effectively communicated to potential stakeholders:

- Addressing a hot topic of significant importance and public interest.
- Highlighting the international dimension of the BIAS project.
- Emphasizing the learning opportunity for target categories to stay up to date.
- Stressing the value of their input in the co-creation process to build fair and trustworthy technology.

Despite several stakeholders expressing interest in attending the workshops, BIAS partners encountered challenges in engaging participants due to various reasons, including:

- Conducting workshops during working hours without compensation.
- Scheduling the workshops during the summer period.
- Some invited participants feeling they lacked the necessary skills to participate, especially those from categories other than HR professionals and AI experts.
- Last-minute dropouts due to unexpected work-related issues and illnesses.

The causes mentioned above, identified by most of our partners (except HI), resulted in non-compliance with the Key Performance Indicator (KPI) of having 24 participants in each workshop. Partners who couldn't meet the KPI during the first round of workshops committed to involving more participants in the second round of co-creation workshops, with the goal of engaging a total of 35 individuals across the two workshops.

The graph below illustrates the composition of stakeholders involved in terms of target categories.



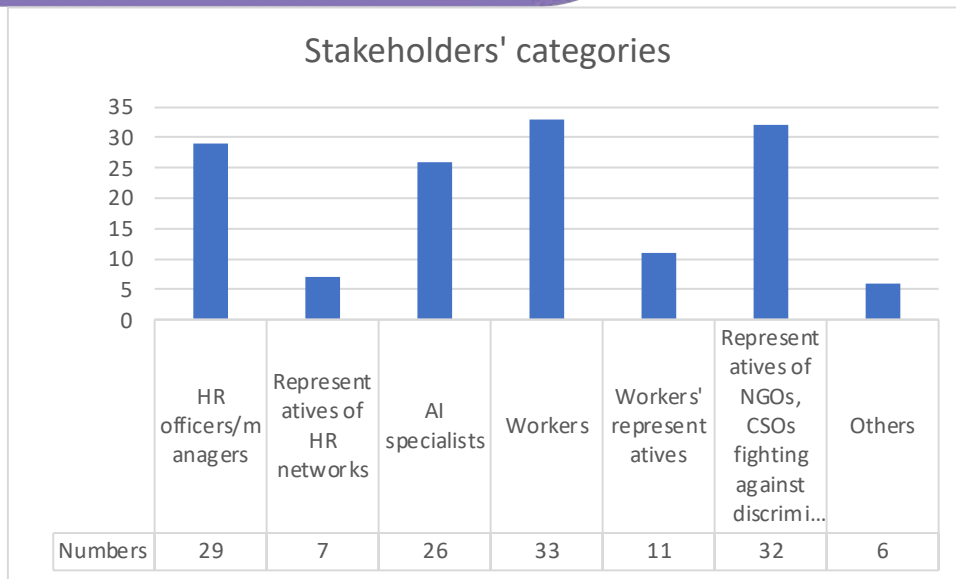


Figure 1 Stakeholders per category in 1st co-creation workshop

As visible and as expected according to the methodology, the most crowded categories are HR officers/managers (29 people + 7 people representing HR networks), workers (33), representatives of NGOs and CSOs fighting against discriminations (32) and AI specialists (26). The “others” category includes stakeholders interested and active in the project’s domains under different respects, for instance representatives of industrial employers’ associations, regional welfare/inclusion policy implementers, researchers from sister projects and academics. In terms of gender representation, the following graph shows that the rate of female participants is of 66,67% reaching and overcoming the set KPI of 45% included in the Education Action Plan 2021-2027.

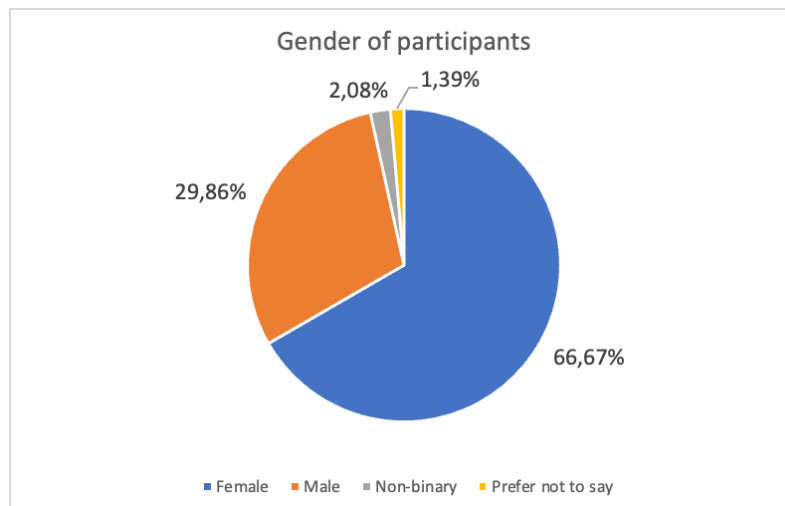


Figure 2 Gender of participants in 1st co-creation workshop

8.2 Overall implementation of the workshops

The reports of the 7 workshops prepared by the involved BIAS partners were made available to WP3 tasks leaders. All partners reported the workshops being successful and meeting the expected results both in terms of the initial panel discussion and the group works, with stakeholders expressing high levels of satisfaction on the experience.





As far as the **initial discussion** is concerned, most partners organized it as a plenary discussion and therefore posing the set questions to all participants (SVEN, NTNU, ULEID, DIGI, BHF, HI), while FARPLAS identified a few panellists to initiate the discussion over the different topics and then left the floor to other participants that wanted to share their opinions and experiences. ULEID organized the discussion in four groups to provide more room to participants to interact and share their thoughts.

Some partners did not manage to discuss all the five topics proposed in the methodology and decided to focus on two-three of them (SVEN, HI). NTNU adapted most of the questions in order to further discuss the issue of “fairness”.

Overall, the discussions were very engaging, and all partners reported that many participants shared their ideas and points of view, although due to the lack of time facilitators in some cases had to interrupt the discussions and move to the following questions. The inputs coming from the discussions are very rich and varied and are summarized in each workshop’s paragraph.

Although being in many cases country specific, some ideas/points of view can be found in the majority of the reports. Many participants shared the idea that AI applications are not neutral for now, since they transfer human bias, but they have the potential of reducing errors if properly trained. Also, the lack of competences on the use of AI applications by HR officers and companies in general was pointed out, together with the importance of involving a wide range of stakeholders in the AI development.

In the discussion on “fairness”, in general participants shared the opinion that the notion of “fairness” is highly context dependent, and it is difficult to assess without knowing the hiring context (DIGI, ULEID). In some cases, “fairness” was associated with “diversity” and “non-discrimination” (ULEID).

Regarding the **group works**, partners overall reported being successful, even though some of them highlighted (mainly during project meetings) some time constraints and the need to shortening the final plenary.

The results of the different activities of the workshop are summarised in [each workshop’s paragraph](#), however an overview of the scenarios and discrimination grounds addressed by partners in the different workshops is presented below.

Three partners adopted and slightly adapted two of the scenarios proposed in the methodology. In particular, two partners (DIGI and BHF) using the tech company looking for a software engineer, while one partner (ULEID) using the research centre hiring a researcher.

The other partners chose other scenarios/job offers better fitting their national/local context. SVEN opted for a job offer in tourism, NTNU one in grocery, HI in health and FARPLAS in automotive.

As far as the grounds of discriminations addressed through the personas’ profiles, all partners adopted gender as one of the dimensions. Only two partners (SVEN, ULEID) used a non-binary approach. Six out of seven partners also adopted race/ethnicity as second dimension, two partners (FARPLAS, HI) sexual orientation. ULEID also included disability as a ground of discrimination besides gender and race/ethnicity.

Given the variety of profiles and the related intersectional dimensions in use, results from the workshops cannot be analysed and compared in a comprehensive way, but it was possible to identify some recurring patterns in terms of gender and race/ethnicity bias.

The figure below lists the main types of bias that were identified in the workshops (in darker colour the most represented ones).



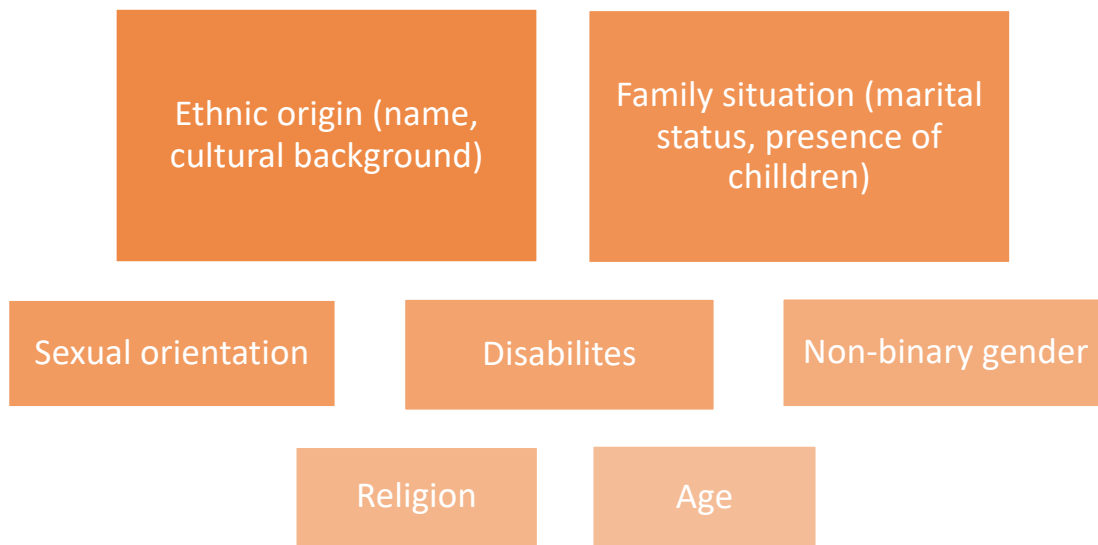


Figure 3 Main types of bias identified in first co-creation workshops

A recurring bias identified in all workshops pertained to the ethnic origin of candidates. In many instances, job offers required a good knowledge of the local language of the country. Consequently, many participants highlighted how the name, origin, and even the picture of candidates from diverse migrant backgrounds could introduce bias into the selection process.

Furthermore, the family situation, including marital status and the presence of children, was often seen as a potential source of bias, particularly against female candidates. Participants discussed how HR specialists/officers might question whether a candidate, especially female ones, would be able to allocate sufficient time to the job. It was also noted that women without children might raise concerns about the possibility of them getting pregnant. However, similar concerns were not typically expressed regarding male candidates with children, as it was often assumed that their female partners would be the primary caregivers. In some cases, having children was seen as a sign of responsibility and commitment for male candidates.

In workshops where these dimensions were addressed (SVEN, ULEID, FARPLAS), potential bias related to disabilities, sexual orientation, the non-binary gender of candidates, religion and age was well noted.

Differences among the different groups and workshops are partly country-related and partly related to the different job offers in use.

Country-related aspects highlighted in the reports include:

- In Italy, small to medium-sized businesses prevail with short decision-making chains. Recruiters are often CEOs, which can lead to a limited multiperspective and professional approach to HR selection and hiring processes. Widespread conservative attitudes and the promotion of "traditional family" representations tend to perpetuate stereotypes against women, migrants, and non-CIS/heterosexual individuals.
- In Türkiye, it is common to associate activism in organizations dealing with gender issues with being gay. The country also experiences political and religious discriminations, particularly towards women wearing headscarves who may face limitations in terms of travel and socializing with colleagues and clients.





- In Estonia, which boasts the highest number of unicorns per capita in the world, many companies require long working hours. As a result, women and individuals with strong family ties may face discrimination.
- In Switzerland, Iceland, Norway, and the Netherlands, a noticeable tendency to discriminate against foreigners was observed. In Switzerland, individuals from ex-Yugoslavian countries often experience discrimination.

8.3 Summary of the reports

8.3.1 1st co-creation workshop in Italy

Introduction

Smart Venice organized the first co-creation workshop on the 12th of July 2023 in Venice.



Figure 4 1st co-creation workshop at SVEN

17 participants took part. Eight more people had registered and had last minute issues that forced them to cancel their participation. Participants represented the following categories.

Table 8 Categories and numbers of stakeholders at SVEN's first co-creation workshop

Categories	Number
HR officers/managers	3
Representatives of HR networks/consultants to HR	2
AI specialists	4
Workers' representatives	1
Representatives of NGOs, networks, organisations fighting against discriminations	4
Other	3

Regarding the represented NGOs, two of them work on contrasting racism and supporting migrants' integration, and one works on gender equality issues.

In the "Others" category we find a representative of an industrial employers' association and their training agency, an Officer from a Regional Authority in charge of projects for the socio-economic integration of migrants and refugees and an academic, a sociologist from the AEQUITAS sister project.

In terms of gender balance, 12 participants self-identified as women and five as men.

Plenary discussion

The discussion was structured selecting few of the five proposed questions, to give enough room for interaction. In particular: question one on the status of Gender Equality, DEI and inclusion policies in





Italian workplaces, question two on pros and cons of the use of AI in HR, and question three on a fair use of AI in HR and its meanings and ways forward.

A good level of engagement was present in the discussion, with no particular dynamics/tensions being noted at this stage.

The main points of discussion can be summarized as follows:

3. Question 1: based on your knowledge and experience, do we have good reasons to be optimistic or should we rather be concerned with the situation in our country?
 - ➔ The geographic component has a significant impact, and differences between Northern and Southern Italy were emphasized. Additionally, participants reported a notable contrast between small and large companies concerning the recognition of diversity, equity, and inclusion policies. They noted that discrimination occurs during various recruitment phases, beginning with the content of job offers.
4. Question 2: what is your opinion and/or experience as far as the use of AI systems in recruitment and Human resources management in general? Are the PROs overly tech-innovation enthusiasts? What points of attention would you advise to balance the CONs?
 - ➔ The lack on competences on the use of AI applications in firms has been pointed out together with the risk of standardisation that the AI applications bring. Two main issues were mentioned: a) how to input the knowledge we already have in the AI based system, as it is not sure we have instruments/tools that allow us to do so and the right knowledge as well; b) how to predict what we can extract from the AI. A limitation is that the technology is still "too young" at present. It was stressed that AI is a great instrument but there is the risk that companies are not ready yet: training and awareness raising activities should be fostered. Also, it was noted that in HR, gender inequalities are the most prevalent.
5. Question 3: what is your view on this? How participation of workers and social partners but also civil society organizations representing minorities can contribute to influence and oversee the use of AI in recruitment and make it fairer?
 - ➔ A general issue concerning inclusive language in working contexts was highlighted. Therefore, CSOs could be engaged in training and awareness-raising activities on this matter.

Group work

Four groups were created, each one having at least an HR professional and an AI expert.

A dedicated scenario was developed from the **hospitality sector** given the prominence of tourism in the local economy. The job offer pertains to a company in the hospitality industry that is seeking a receptionist/manager for a hotel in Venice. The person will be responsible for managing a guesthouse with 11 rooms in Venice, working five days per week, with full phone and in-person availability. The main responsibilities include welcoming clients, promptly responding to their messages, and assisting them in enjoying the city. Additionally, the role involves managing the cleaning activities. The ideal candidate should be fluent in both English and Italian, and ideally, also in Spanish and French. They should possess a dynamic attitude and strong problem-solving skills. The person is expected to work five days a week, including Saturdays and Sundays. Initially, a short-term contract of 3 months is offered with the possibility of converting it into a long-term contract.

Gender and race/ethnicity were chosen as factors for developing the four personas, with a non-binary approach for gender:

- Persona Group 1: A white, young man born in Italy, cisgender, single, with six years of experience in similar positions, holding a master's degree in international management and a master's





course in tourism management. He has a strong command of English, Spanish, and French. His hobbies include yoga and cooking, and he is portrayed as good-looking and elegant in his picture.

- Persona Group 2: A white man, transgender, married with a child, possessing 20 years of experience in the tourism sector, with a diploma (no degree) in Tourism economics and management. He has studied Spanish and French, loves animals, teaches English and Spanish to children, and describes himself as dynamic and well-organized. He appears casually dressed in his picture.
- Persona Group 3: An Asian man born in Italy, cisgender, married with three children, having 20 years of experience in the hospitality sector and as a receptionist. He holds a master's degree in Tourism management from Delhi. He is interested in chess, has full availability, pays attention to details, and has very good knowledge of English and Italian, as well as proficiency in another foreign language. He presents himself as elegant and professional in his picture.
- Persona Group 4: An Asian young man, transgender, single, with eight years of experience in the hospitality sector, a diploma (no degree) in tourism, a passion for badminton, and volunteer work with an NGO focused on human rights. He describes himself as having excellent organizational skills and proficiency in five languages (including English, Spanish, and French).

The group work proceeded smoothly, with no issues arising. There was a high level of dialogue and collaboration among the groups. At times, the AI specialists appeared to be slightly less at ease than others, as their background and professional experience were less applicable to a simulation that did not involve technology.

Discussions related to potential biases that candidates might encounter were among the most productive in terms of identifying potentially biased words, sentences, and word associations. However, the writing of the cover letter was somewhat constrained in terms of time, resulting in relatively short letters. Mitigation of biased words from the cover letter typically focused on a couple of specific words.

The most significant findings and dynamics that emerged during the various activities can be summarized as follows:

During the **first activity** potential biases and stereotypes related to the job offer were discussed, including potential positive bias towards male candidates and potential negative bias towards candidates with family responsibilities or migrants. The groups also examined potential discriminatory aspects of the job offer, such as a potential mismatch between responsibilities and compensation and the expectation of complete availability, including for working long hours and during weekends.

Each group highlighted different aspects of the job offer, such as the demanding requirements and the potential targeting of young, single female candidates in some cases, or conversely, male candidates to ensure they are free from caregiving duties.

It was also noted that the managerial component of the tasks made it less suitable for a junior profile. For some, a potential bias would be towards non-native/non-Italian, particularly migrant candidates, as fluency in Italian is required, and knowledge of the place and its cultural heritage would be desirable.

During the **second activity** several discussions arose in the groups:

- Group 1: Divergent opinions on whether to include complimentary comments about the position or the employer in the cover letter, as well as the tone and extent to which quantifying the candidate's previous results would be useful. Key words for the cover letter, such as "multitasking" and "able to work under pressure/stress," were considered potential sources of bias.
- Group 2: Discussion on how the candidate's place of birth or current location might impact their knowledge of Venice and on the need for commuting, which might lead to a negative bias for the





specific position. The main negative bias identified pertains to the candidate's transgender identity, while many positive biases are listed.

- Group 3: Emphasis on the candidate's extensive experience and qualifications, leading to potential positive and negative biases. Reflection on potential prejudices related to being a foreigner. The group decides that information on ethnic background/nationality should not be included in the cover letter, nor should information about family status.
- Group 4: Identified biases related to various aspects, including the candidate's gender, ethnicity, education, work experiences, and image. The Asian origin can lead to both positive and negative stereotypes, such as being hardworking and organized or being perceived as a temporary resident due to youth and foreignness, potentially making them seem unstable. Biases regarding transgender identity are also discussed. Regarding the candidate's image from the picture, biases are associated with Asian features, tattoos, long hair, and attire.

The discussion over the cover letter (**activity 3**) led to the following considerations:

- Group 1: Certain terms, such as "multitasking," when associated with a male, may not be perceived as credible by HR. Similarly, certain words and phrases that refer to the candidate as male could create positive biases.
- Group 2: Identified biases include gender and age bias. The candidate's background, specifically the reference to a multicultural context, could create doubts and negative stereotypes if the recruiters are specifically looking for an Italian candidate. The discussion also addressed the topic of the candidate's transgender identity, with participants suggesting different approaches to handle it. Certain words and phrases in the cover letter were identified as potential sources of bias, such as "twenty years of experience" being associated with being old and "multicultural context" raising questions about the candidate's origin, potentially leading to negative stereotypes.
- Group 3: The group recognized the theme of the candidate's need for work and debates whether to include information about soft skills. Risks of bias and prejudice were identified, primarily related to the candidate's name "Ahmed," the degree obtained in Delhi, and the professional background in housekeeping. Potential biases originating from the candidate's image were discussed, with perceptions of the candidate being overqualified for the position. Moreover, mentioning having three children plus being a migrant could lead to identification with a person in need.
- Group 4: Biases were related to various aspects, including the candidate's gender, ethnicity, education, work experiences, and image. The Asian origin could lead to both positive and negative stereotypes, such as being hardworking and organized, or being perceived as a temporary resident due to youth and foreignness, potentially making them seem unstable. Biases regarding transgender identity were discussed. Regarding the candidate's image from the picture, biases are associated with Asian features, tattoos, long hair, and attire.

The results of the activity 4, as well as all bias words/sentences identified are reported in the spreadsheet made available to WP3 tasks leaders.

Further interesting highlights from the workshop

Additional noteworthy highlights from the workshop include the tendency of some HR profiles to self-identify with over-demanding/suspicious conservative employers, emphasizing and exaggerating potential negative and positive biases. This may be explained by participants considering the local industrial context, characterized by SMEs run by conservative employers and having short decision-making chains and limited or absent HR functions. In such a context, featured by certain conservative attitudes that tend to reproduce stereotypes, achieving "fairness" is considered challenging and CSOs tended to highlight negative bias against minorities, particularly migrants, with whom at times other stakeholders did not fully agree.





Key takeaways from the Italian co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 9 Key takeaways from the first Italian co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Words/sentences that fostered discussion
Gender of the candidate	Positive bias towards male candidates for job offer requirements Positive and negative bias towards female candidates (disagreement among the groups) Negative bias emerged towards the transgender candidate	
Family situation	Negative bias towards candidates with family responsibilities and especially women	
Migrant background of the candidate	Negative bias towards migrants -> fluency in Italian is requested Negative bias related to the name of candidates having migrant background Positive bias in case of Asian origin -> hardworking and organized	“Ahmed”
Age of the candidate	Positive bias towards young candidates -> demanding requirements Negative bias towards young candidates -> managerial skills required	
Previous working experience of the candidate	Positive and negative bias depending on the association with the age of the candidate	“twenty years of experience”, “multicultural context” “overqualified”
Soft skills	Positive and negative bias -> discussion on including or not in the cover letter	“multitasking”, “able to work under pressure/stress”





8.3.2 1st co-creation workshop in Norway

Introduction

NTNU organized the first co-creation workshop on the 9th of June 2023 in Trondheim.



Figure 5 1st co-creation workshop at NTNU

24 people initially registered to attend, of these, two did not come, however, one person that had not registered came at the last minute bringing the number of total participants to 23. Participants represented the following categories:

Table 10 Categories and numbers of stakeholders at NTNU's first co-creation workshop

Categories	Number
HR officers/managers	7
Representatives of HR networks	0
AI specialists	3
Workers	6
Workers' representatives	2
Representatives of NGOs, networks, organisations fighting against discriminations	5

Representatives of NGOs came from one organization empowering immigrants, one human rights organization, one organization targeting marginalized communities in the Middle East / North America and one organization promoting LGBTQIA+ rights. One of the organizations' representatives was also an AI expert.

In terms of gender balance, 13 participants self-identified as women and 6 as men and 2 identified as non-binary, while one preferred not to say.

Plenary discussion

The discussion was structured by selecting five questions, one from those proposed in the methodology, and four original questions around the theme of fairness in recruitment and challenges to address. Although the discussion started a little slowly, participants became actively engaged, addressing general comments on the state of bias in recruitment in Norway based on their experiences. Some tension emerged when one participant from academia and another from the private sector disagreed on the importance of using or not using names when applying for jobs.





The main points of discussion can be summarized here below:

1. Question 1: What are your thoughts on employment and recruitment in Norwegian context? What is the most important issue that needs to be solved right now?
 - ➔ Blind review was proposed as solution for discrimination in recruitment, although not considered very realistic.
2. What does fairness mean when it comes to recruitment? What keywords, in both Norwegian and English, would indicate fairness?
 - ➔ Basing the selection solely on skills and background was suggested as an objective and equitable approach.
3. How important is it to achieve fairness during recruitment? Is it the basic rule or just a bonus?
 - ➔ No inputs were provided on this question.
4. What do you care most during recruitment as a worker/HR practitioner/organization member?
 - ➔ The discussion focused on the recruitment process in an academic context and the potential use of blind CVs. The final considerations revolved around the role of technology in making the process fairer, acknowledging the differences across industries and sectors.
5. Do we have good reasons to be optimistic or should we rather be concerned with the situation of equality, diversity, and inclusion (EDI) in Norway?
 - ➔ The tendency to exclude PhD students from certain countries, which happens in both the academic and private sectors, was highlighted, particularly in the university sector.

Group work

Four groups were created, each consisting of at least an HR professional, an AI expert, and a worker.

The scenario chosen by NTNU was a **grocery store chain** looking for a cashier. The job is part-time for the checkout area of the supermarket, with working hours between 16:00-21:00. However, the time shifts may change, and there is likely to be more shifts and work during holidays. The main tasks for the job include checking out items, taking payments, and making the customer feel welcome. It also requires working as a cashier in the self-service checkout, picking up baskets, cleaning, filling goods, and helping customers. The ideal candidate should be able to speak a good level of Norwegian and have social skills to interact with the public. They are also required to have the ability to handle stress.

Gender and race/ethnicity were selected as grounds of discrimination for developing the four personas:

- Persona Group 1: A white young woman born in Norway, cis, married with no children, with five years of experience in design. She has a high school diploma and attended a design summer school. Her hobbies include skiing and hiking.
- Persona Group 2: A white middle-aged man, born in Norway, cis, married with three kids. He has ten years of experience working as a carpenter, with education from a vocational school in craftsmanship and design. He is a Norwegian native speaker and fluent in English. He enjoys hiking and outdoor activities and has a positive personality.
- Persona Group 3: An Iranian young man, cis, born in Iran, married with three children. He has five years of experience as a delivery driver and four years of experience as a service worker in a cafe. He holds a high school diploma from Tehran and attended a Norwegian language school. He is proficient in Persian, English, and fluent in Norwegian.
- Persona Group 4: A Lebanese young woman, cis, single. She has a bachelor's degree from Lebanon and a master's degree in Norway. She has working experience in business development and marketing. She speaks Arabic as her native language, is fluent in English, and has basic knowledge of Norwegian. She enjoys traveling and doing yoga and appears friendly in her photo.





The group work proceeded smoothly, with discussions and co-creation in each group progressing well. HR experts provided their opinions on the profiles and suggested how to write the cover letter, which was generally followed by other stakeholders. In general, all participants contributed, including through the use of sticky notes, except for one group in which one participant was more active than the others. The writing of the cover letter was the most challenging activity to implement since people do not always enjoy writing cover letters for themselves, making it somewhat tedious to do so for the personas.

The most relevant contents and dynamics that emerged in the different activities can be summarized as follows.

During the **first activity** the discussion in each group about potential biases focused on the following topics: level of the job, request for a picture, language requirements, request for “thrive in a fast-paced environment”, and late working hours. All groups agreed that the position was considered a “low-level” job and that the working hours could be a disadvantage for applicants having children or other caregiving duties in their personal life. The requirement for a good level of Norwegian was also considered a potential bias. Asking for a personal photo was also considered a potential bias. Finally, the person’s mental health condition could also generate a bias.

During the **second activity**, discussions in the different groups concerned:

- Group 1: The overqualification of the candidate, which can be considered both positive and negative. Also, her identity as an Iranian woman who is single is considered as a positive bias by all group members.
- Group 2: The major point of discussion was on whether it would be wise or not to mention the family situation, since the job requires working unusual hours. However, it was considered positive to mention it because it would show his responsibility and ability to manage three children, assuming that their mother takes care of them while he is at work.
- Group 3: The candidate is a married woman in her thirties, without children, and participants discussed whether or not that would imply that she might get pregnant in the near future and the potential impact on the job.
- Group 4: "Iranian" was considered as a positive bias based on the stereotype that Iranian people are good at selling, although some participants were more concerned about the prejudice related to nationality due to the implied lack of Norwegian skills. Another discussion was around the information on the family situation, which is still related to working hours.

The discussion over the cover letter (**activity 3**) led to the following considerations:

- Group 1: The group discussed what the focus of the cover letter should be. They had different opinions on whether it should highlight the candidate's identity as a foreigner.
- Group 2: A discussion arose about why a person of this age with kids would want a job like this. Based on the picture, the candidate seems very responsible and serious, but may not be a good communicator.
- Group 3: The group decided to remove the word "woman" from the initial part of the cover letter in order not to disclose potential pregnancy.
- Group 4: The education level of the applicant was considered a negative bias due to overqualification. Biases related to extra working hours were seen by group members as the biggest risk to address. Some members argued that mentioning the term "cisgender" explicitly signals solidarity, but at the same time, the term can be perceived as political and potentially backfire on the applicants.

The results of the activity 4, as well as all bias words/sentences identified are reported in the spreadsheet made available to WP3 tasks leaders.





Further interesting highlights from the workshop

Some HRs expressed personal preferences regarding certain personal attitudes/personalities of the applicants. As for country-specific biases, the level of the Norwegian language and citizenship were identified.

Key takeaways from the Norwegian co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 11 Key takeaways from first Norwegian co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Gender of the candidate	Positive bias towards female and single Iranian candidate (intersectional bias)	“woman” “cisgender”
Family situation	Negative bias towards candidates with family responsibilities due to working hours Positive bias towards female candidate with family -> responsibility and management attitudes Potential negative bias towards a married woman in her thirties -> possible children	“married with three children”
Migrant background of the candidate	Negative bias towards non-Norwegian candidates since a good level of local language was required Positive bias towards Iranian candidate -> “they are good at selling”	
Previous experience of the candidate	Positive and negative bias towards the overqualification of candidate	“overqualified” “able to join the labour market”
Request for a picture	Potentially leading to bias	

8.3.3 1st co-creation workshop in The Netherlands

Introduction

ULEID organized the first co-creation workshop on 4th of July 2023 in Leiden.



Figure 6 1st co-creation workshop at ULEID





30 people initially registered, but 20 attended. Participants represented the following categories:

Table 12 Categories and numbers of stakeholders at ULEID's first co-creation workshop

Categories	Number
HR officers/managers	5
AI specialists	2
Workers	4
Representatives of NGOs, networks, organisations fighting against discriminations	6
Other	3

Represented NGOs operate in the following areas: human rights, disability, race & technology, privacy issues and research. In the “other” section, academics of the faculties of economics and law were present.

In terms of gender balance, 13 participants self-identified as women and 6 as men and 1 identified as non-binary.

Plenary discussion

To facilitate the discussion, participants were divided into four groups, each dedicated to exploring the five proposed questions. Each question was introduced with a piece of news to help participants contextualize it. One moderator and a note-taker were present in each group. Participants were highly engaged in the conversation, although it was common for them to shift the focus towards their own professional and sometimes personal experiences.

The main points of discussion can be summarized here below:

1. Question 1: based on your knowledge and experience, do we have good reasons to be optimistic or should we rather be concerned with the situation in our country?
 - ➔ Technology is not neutral; currently, it transfers human bias into it. The production and use of technology reflect a long-lasting asymmetry of power, excluding certain social communities, such as people with disabilities. Technology lacks a 'human touch' and empathy. However, some participants express optimism.
2. Question 2: what is your opinion and/or experience as far as the use of AI systems in recruitment and Human resources management in general? Are the PROs overly tech-innovation enthusiasts? What points of attention would you advise to balance the CONs?
 - ➔ Some participants treat AI applications for HR purposes with caution. They suggest that AI applications should primarily facilitate job applicants rather than employers. A con identified is the technological divide that the use of AI applications could imply, favouring big companies. Among the pros, gender-neutral language, avoidance of repetitive tasks, and the creation of templates are noted. AI can also foster the creative process.
3. Question 3: what is your view on the role that AI based technology can play to favour or to hamper EDI in hiring processes in particular?
 - ➔ Most participants argue that AI can hamper EDI in the selection and recruitment process, as it is likely to reproduce existing diversity bias, and create new ones. However, when responsibly used and with proper training, AI can also promote EDI.
4. Question 4: how would you define it and to what extent such definition is context dependent in your view?
 - ➔ All participants agree that fairness is contextual and dynamic. Fairness often overlaps with non-discrimination but is more nuanced in practice. Different views emerged, but it was stated that fairness in recruitment could correspond to diversity-oriented HR practices.





5. Question 5: what is your view on this? How participation of workers and social partners but also civil society organizations representing minorities can contribute to influence and oversee the use of AI in recruitment and make it fairer?

→ The involvement of a diverse pool of stakeholders should occur from the design stage till the regular maintenance of technology, even though it is considered difficult to concretely adopt a co-creation approach.

Group work

Four groups were created, each one having at least an HR professional.

The scenario adopted by ULEID was slightly adapted from one of the proposed ones. It covered a job application for a **post-doc researcher in bio-medical engineering**. Applicants were expected to develop an ambitious project for their future group and to contribute to the center's strategy based on excellent science, internationalization, translation and talent. Apart from the outstanding scientific output, the candidates should prove that they are active in applying competitive proposals as principal investigators. Any mobility experience, e.g., a stay in another country/region, was considered a valuable contribution. Leadership and people management; critical judgment in identifying and executing research activities; strategic vision for the future of the research field; income and funding generation; knowledge generation and transfer; collaboration; inclusion; excellent communication and networking; excellent knowledge of the English and Spanish languages were considered desirable skills and competencies.

Gender and race/ethnicity were selected as grounds of discrimination, but and disability was also covered due to the involvement of relevant stakeholders. Gender was examined through non-binary lenses.

- Persona Group 1: A Black young man from Indonesia with a PhD in mechanical engineering from there. Never travelled. Some basic research experience. Hobbies: football, martial arts, and doing nerdy stuff related to the IT world. He wears casual and smiles in the picture.
- Persona Group 2: A White young woman with physical disability from the Netherlands. PhD in mechanical engineering from the best Dutch university in this field. Some work experience in the private sector. Hobbies: accessible hiking, adaptive sports, and technologies, advocating for inclusion. She sits in the wheelchair and looks thoughtful in the picture.
- Persona Group 3: A White young man from Turkey. With a severe physical disability. PhD in mechanical engineering from Turkey. And some working experience there. Hobbies: adaptive sports and technologies, advocating for people with disabilities in STEM. He sits in the wheelchair, wears casual and smiles in the picture.
- Persona Group 4: A White very young transgender man from Mexico with mental disability. PhD in mechanical engineering and some working experience in research and in the private sector from there. Hobbies: Art, music, environmental conservation, and disability advocacy. He looks very young in picture.

The group work during the workshop ran smoothly, allowing everyone to have an opportunity to express themselves and contribute to the discussions. The participation was balanced, ensuring a diverse range of ideas and viewpoints. Each individual's unique perspective influenced their interpretation of the tasks and their problem-solving strategies, adding richness and depth to the discussions.

Among the proposed activities, the drafting and writing process of the cover letter was perceived as slightly more challenging and less smooth to implement. The allocated time for the activity may have been perceived as insufficient, which could have added a sense of pressure and hindered the smooth implementation of the task.

The most relevant contents and dynamics emerged in the different activities can be summarized as follows.





During the **first activity** participants noted the male-oriented nature of the job offer, as well as the intersection between mobility and care duties. Some people questioned the necessity of certain job requirements. Overall, the job application was considered too demanding.

During the **second activity**, it emerged that most groups believed that their job applicant could be the perfect match and tried to have a cover letter covering all the requirements of the job description, rather than favouring certain aspects. Overall, the focus was on the past experiences, background, and skills of the job candidate rather than their personal characteristics. In one group, there was a discussion about the explicit reference to legally protected grounds: from an anti-discrimination law angle, it could be an added value at the intersection with positive discrimination, while data protection law generally prevents this data processing to level differences out.

During the discussion over the cover letter (**activity 3**), all the groups put great emphasis on the possible bias arising from the picture, especially in terms of age. Discussion also related to race. Overall, it appeared that bias is contextual. On the explicit reference to disability, one group argued that it allowed the candidate to make their needs clear.

During **activity 4** rather than focusing on specific words to change, people discussed whether to explicitly refer to legally protected personal characteristics or not.

Further interesting highlights from the workshop

Participants from different categories, such as HR officers, workers, and AI specialists, tended to focus on specific biases and related aspects. These different perspectives added valuable insights and depth to the discussions. While HR professionals showed heightened attention to analysing the pictures included in job applications, scholars and participants with expertise in AI were more adept at scrutinizing the job requirements and expectations, leveraging their past experiences to effectively match the job applications with the necessary qualifications.

In the workshop discussions, certain aspects, topics, and biases emerged that could be identified as country-specific. For example, participants acknowledged that the name and country of origin could lead to subconscious bias favouring candidates from the Netherlands or similar countries, potentially overlooking qualified candidates from other regions.

Another noteworthy aspect that surfaced was the impact of visual cues, particularly in the context of disability. Participants noted that when a person with a disability appears with no physical disabilities in a picture, there was a tendency to assume that accommodating their needs would be relatively easy.

Key takeaways from the Dutch co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 13 Key takeaways from the first Dutch co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Gender of the candidate	Negative bias towards women -> due to the nature of the job offer and the required skills Positive bias towards transgender candidate	"Transgender"
Family situation	Negative bias towards women with children due to the requirement to travel internationally	
Migrant background of the candidate	Negative bias for candidates not having work experience in NL	"Turkish education" "migrant background"





Disability of the candidate	Positive bias towards disabled candidate -> indication of bravery Negative bias towards disabled candidate -> lack of resources and obstacle for travel requirement	“Physical Disability” “Mental disability”
Age of the candidate	Negative bias towards very young candidate	

8.3.4 1st co-creation workshop in Iceland

Introduction

The University of Iceland (HI) organized the first co-creation workshop on the 9th of June 2023 in Reykjavik.

24 participated. Participants represented the following categories:

Table 14 Categories and numbers of stakeholders at HI's first co-creation workshop

Categories	Number
HR officers/managers	4
Representatives of HR networks	2
AI specialists	4
Workers	4
Workers' representatives	4
Representatives of NGOs, networks, organisations fighting against discriminations	6

Represented NGOs operate in the following areas: gender equality, disability, ethics in HI, LGBTQI rights. In terms of gender balance, 15 participants self-identified as women and 9 as men.

Plenary discussion

3 *ad hoc* questions were elaborated from facilitators in order to foster the discussion. Participants were very engaged in the discussion, so facilitators opted for focusing on only two questions rather on all five.

The most interesting points of view emerging from the discussion can be summarized as follows:

- Question 1: What is your opinion on or experience with the use of artificial intelligence in recruitment and human resource management? What do we need to keep in mind to achieve the best results?
 - participants had strong and partly different opinions but not necessarily experience using AI in recruitment. In general, they considered it positive that AI could reduce bias in recruitment. However, they raised concerns that AI could be “fooled”, creating more problems than it solves. Therefore, they stressed the importance of maintaining critical thinking.
- Question 2: What is your view on the role that AI-based technology can play in promoting or hindering issues such as equality, diversity and inclusion?
 - AI has the potential to select the best person for a job, but it may reject candidates with disabilities due to a lack of data on hiring such individuals. Ensuring that people with disabilities are not excluded from AI is crucial.





Group work

Four groups were created, each one having at least an HR professional, an AI professional, a worker or workers' representative.

The scenario adopted by HI was based on a real job offer for a **Head of communication** in a health drinks company. The company uses technology based on clear rock water, fatty acids, and natural minerals for health-related problems. The company operates in an international market and is on the fastest-growing companies in Europe. The main tasks and responsibilities for the job is to formulate and follow a strategy in promotional and web matters, press releases, media relations and social media communications. Education and qualification requirements were formulated as well as the required excellent skills in Icelandic and English. This is a full-time job, which requires considerable travel, both domestically and abroad.

Gender and race/ethnicity were selected as main grounds of discrimination. Also, sexual orientation was added as dimension.

- Persona Group 1: A white man born in Iceland, cis, married, and father of three young sons. He has a master's degree in digital engineering and marketing from Erasmus University in Rotterdam. He has ten years of experience in marketing and public relations jobs. Hobbies include football and computer games. In his spare time, he coaches young boys in soccer. He is good-looking and elegant in the picture.
- Persona Group 2: A white woman born in Iceland, cis, single (divorced), and mother of three children. She has a master's degree in cultural management and has worked with marketing and PR jobs since 2000. Hobbies include choir singing and mental health/self-help.
- Persona Group 3: A young woman, born in Iceland but with parents from Thailand. She is a lesbian, has a female partner, and two children. She has a master's degree in digital marketing from Columbia University, New York, and has worked in marketing for the last five years. She is athletic and competes internationally in triathlons and coaches people for "iron man." She looks sweet in her photo.
- Persona Group 4: A young man from Vietnam, moved to Iceland with his parents when he was eight years old. He is cis, single, with no children. He has a BS degree in computer sciences and works as an expert in software development with a focus on automation. Hobbies include traveling. He looks young and happy in the picture.

The group work during the workshop ran smoothly. The group work was slightly shortened to allow more time for the plenary. Collaboration among the different stakeholders was smooth and balanced.

The most relevant contents and dynamics emerged in the different activities can be summarized as follows.

During the **first activity**, participants pointed out that the advertisement was not suitable for young mothers due to the required travels. It was also discussed that the job, as described in the scenario, might not be suitable for disabled people, such as those who use wheelchairs. The requirement for communication skills was seen as excluding people with autism. The need for the employee to fit into Icelandic culture was seen as a bias against foreigners. The job offer was also found to contain age discrimination.

During the **second activity** discussions in the different groups concerned:

- Group 1 -> The cover letter portrayed the candidate as a "typical Icelandic" ambitious and narcissistic white man. The group emphasized the candidate's qualities and competences. Although he has three children, and the job requires traveling, none of the participants expressed concern about his work-life balance. The group described the candidate as someone who loves





traveling for work to get away from the family situation, taking the support he has from his wife for granted (gender bias).

- Group 2 -> The group did not find it positive to emphasize the interests and hobbies of the candidate in a cover letter, especially mentioning a self-help book, as it could lead to prejudices. The candidate's children's ages were unknown from the profile, and this was seen as a bias against a single mother with children, a bias that would not apply to a single father.
- Group 3 -> The group discussed the competitive personality of the candidate, questioning whether it was good or bad for the job. They wondered if she would transfer her competitive mood to the company. This was considered a gender bias, as the same reflections might not apply to a male candidate. Being very socially active, the group wondered if she would have time to focus on the job. They also questioned why she disclosed being gay, as it was seen as an indirect message to communicate strong support in household activities and managing children.
- Group 4 -> The candidate seemed not having all the required skills for the position. His origins were mentioned in the cover letter, stressing that they could represent a bias.

The discussion over the cover letter (**activity 3**) led to the following considerations:

- Group 1 -> the cover letter presented the candidate as a self-confidence person, which could have been seen both positive and negative, since he might look arrogant. The fact of looking extravagant in the description of his experiences was perceived as more typical for men than women. The group member agreed that the cover letter gave a picture of the candidate as a "typical Icelandic" ambitious man with a lot of support from home (gender bias);
- Group 2 -> the discussion arose around the fact that the candidate mentions her "children" in the cover letter which can result in a bias against her. Another bias was associated to the fact that she communicated being a "single mother".
- Group 3 -> the group identified a possible bias in the age of the candidate, since she is young which could lead to negative bias but she is experienced which is positive.
- Group 4 -> The candidate appeared to be privileged. The group observed that the candidate could have emphasized that he had managed to overcome different obstacles as an Asian immigrant in Iceland. They also observed that the candidate gave a strong image of himself as a "tech geek," which might raise doubts about his social skills. The candidate might be discriminated against due to his Asian appearance and young age.

During **activity 4** rather than focusing on specific words to change, the groups discussed on improvements that candidates could make in their cover letters.

- Group 2 -> The group suggested that the candidate should emphasize that being a mother of three children does not prevent her from traveling for work and that she is well organized. They recommended removing the fact that she wrote a "self-help" book, as it could be counterproductive.
- Group 3 -> The discussion revolved around whether a 28-year-old and athletic woman could be seen as a reliable person for the position. The group suggested writing something that counteracts potential biases.
- Group 4 -> The group rephrased the cover letter, moving the focus from presenting the candidate as a "tech geek" and an immigrant who likes traveling to emphasizing his experience and the challenges he faced as an immigrant.

Further interesting highlights from the workshop

Participants from different categories tended to focus on specific biases, especially the representatives of NGOs, who elaborated on the discriminations their members experienced.

During the plenary, it was stressed how the male candidate, having three children, did not explain in the cover letter how he would deal with his family situation (since the job required traveling). On the other





hand, the cover letters of female applicants had to explain how they managed family situations, which was considered a gender bias.

A country-specific aspect that emerged was related to language barriers. Being Iceland a small country, people feel the need to preserve the local language, which represents a bias against foreign individuals.

Key takeaways from the Icelandic co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 15 Key takeaways from the Icelandic workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Gender of the candidate	Positive bias towards male candidates (even if having children) Negative bias towards female candidates also connected with emerging competitive personalities	“competitive” “female gay relationship”
Family situation	Negative bias towards young mothers due to the requirement to travel but positive bias showing responsibility and management skills	“single mother” “no kids”
Migrant background of the candidate	Negative bias towards foreign candidates -> request for Icelandic culture	“Asian”
Disability of the candidate	Negative bias towards disable candidate -> using a wheelchair could have represented a problem within the job. Also having autism would not fit the job offer.	
Age of the candidate	Negative bias towards young candidate	“young”
Hobbies	Negative bias from hobbies -> hobbies not to be mentioned in the cover letter and CV	“choir as therapy”

8.3.5 1st co-creation workshop in Turkey

Introduction

Farplas organized the first co-creation workshop on the 18th of June 2023 in Gebze.



Figure 7 1st co-creation workshop at FARPLAS

18 people participated. Participants represented the following categories:





Table 16 Categories and numbers of stakeholders at FARPLAS' first co-creation workshop

Categories	Number
HR officers/managers	3
AI specialists	3
Workers	7
Representatives of NGOs, networks, organisations fighting against discriminations	5

Represented NGOs operate in the following areas: women's rights and disability. In terms of gender balance, 12 participants self-identified as women and 6 as men.

Plenary discussion

Four questions were elaborated from the ones proposed in the methodology and asked to four previously identified panelists. The other participants listened to the speakers and were invited to add further comments/insights. The questions posed and the main points of discussion are summarized below:

1. Question 1: The first question was posed to the CEO of an automotive company and was introduced by a statement regarding inclusion and diversity in working life in Turkey, according to a study conducted by Deloitte that was aimed at exploring whether participants in the research have ever been discriminated against in their working life. The panellist was asked: *Based on your knowledge and experience, do we have good reasons to be optimistic or should we be more concerned with the situation in our country?*
 - ➔ The panellist stated that managers can be biased. A significant generation gap can trigger the formation of bias. Different cultures have different perceptions of prejudices; therefore, it is essential for the recruiter to understand the candidate's culture, and for the candidate to understand the institution's profile.
2. Question 2: the second question was posed to a senior partner at a global management consulting firm. The panellist is also an advisor on diversity and inclusion issues. *What is your opinion and/or experience on the use of artificial intelligence systems in recruitment and Human resources management in general? Are PROs extreme tech-innovation geeks? What points would you recommend paying attention to balance the CONs?*
 - ➔ The panellist believed that AI is very reliable in recruitment if fed with correct data, and it can reduce the margins of errors to nearly zero.
4. Question 3: the third question was posed to an experienced partner with a long experience in recruitment. She also provides consultancy to companies on recruitment. *What is your view on the role AI-based technology can play in supporting or preventing EDI, especially in recruitment processes?*
 - ➔ The importance of ensuring that HR specialists can use AI technologies correctly was stressed. Training of HR staff will be crucial to make AI contribute to the recruitment process.
5. Question 4: the fourth question was posed to a senior talent acquisition who has also disabilities and activist in a civil society organization. *What is your opinion on this issue? How can the involvement of workers and social partners, as well as non-governmental organizations representing minorities, contribute to influencing and monitoring the use of AI in recruitment and making it more equitable?"*
 - ➔ In her opinion, barriers generated by the fact of using a wheelchair do not have an impact on recruitment. Inclusive policies and strategies followed by institutions have had a positive impact on this and have triggered more inclusion of disabled people in the





business world. However, since there are still issues to be faced, minority representatives should be very active and have their say in technological development.

Group work

Four mixed groups were created, with four or five people each.

The scenario adopted by FARPLAS concerned a company in the automotive sector looking for a **project and sales engineer**. The candidate is expected to take an active role in identifying new customers, determining new commercialization strategies and following the developments in the automotive sector. For this role, being aware of the new innovative and technological solutions and trends is crucial. The company is implementing projects in German, French and Korean, in that sense language ability especially English is essential. Candidates must have five-seven years of experience in a similar position and in the relevant sectors. Computational skills and quality system applications are highly important. The applicants should have a degree in Engineering. Applicants are expected not to have any travel restrictions.

The personas chosen took into consideration common prejudices that HR people have in Turkey regarding gender and sexual orientation as grounds of discrimination:

- Persona Group 1: A White, Turkish, cis man, married and father of 2 children. He is 43 years old with a degree in mechanical engineering and a master's degree in economics and finance. His hobbies include sailing, diving, and chess.
- Persona Group 2: A Turkish, White, cis and single man. He is 31 years old, with a degree in industrial engineering, advanced English skills, and experience in project engineering. His hobbies include cycling, drawing, pottery, yoga, and professional photography. He actively participates in cultural and gender-focused civil society organizations.
- Persona Group 3: A Turkish, cis woman, 29 years old and recently married. She has experience in relevant sectors as a project and sales engineer integrated with technology transformation. Her hobbies include traveling, Pilates, and tennis, and she volunteers at animal shelters.
- Persona Group 4: A Turkish, cis woman, married with 3 children and wearing a headscarf. She graduated from the material science and engineering department and has lived abroad for over 2 years. Her hobbies include gastronomy and walking.

The group work during the workshop ran smoothly, and all participants contributed regardless of their specific field of expertise. The activity of elaborating the cover letter was perceived as challenging, and some groups did not write a fully structured cover letter.

During the **first activity**, some participants pointed out that hiring a person with disability would have been challenging due to the frequent travels required for the position. The position was seen as suitable for both genders. Possible biases related to expected knowledge of foreign languages, the use of technologies, and the potential unsuitability of candidates with economics and business backgrounds were discussed. Expected skills, communication, managerial and leadership competences, extensive experiences abroad, as well as in the R&D field were identified.

During the **second activity** the groups elaborated the cover letters. The main highlights are as follows:

- Group 1 -> The cover letter explained the reason why the candidate wanted to return to corporate life after working as an independent consultant. It also addressed the candidate's absence of travel barriers, despite having two children and explained the reasons for taking six years to graduate rather than the usual five. The mention that the candidate was a "commander" while doing military service was considered positively.
- Group 2 -> The cover letter mentioned the absence of military service, and the group debated whether this information should be added or not. The candidate's experiences abroad were perceived as an indication that the candidate did not find what he expected in Turkey, and therefore, he might leave the country and the company. The cover letter emphasized the





candidate's commitment to the company. The group discussed the candidate's participation in gender and culture-oriented non-governmental organizations, which raised concerns about him being perceived as gay, even though this information was not provided in the persona's profile. Participants associated activism in such organizations with the need to be "heard" in society and representing a minority. The fact that the person is single also supported this idea.

- Group 3 -> The group mentioned in the cover letter that the candidate is results-oriented and explained how she would align her hobbies with her working life.
- Group 4 -> The group explained in the cover letter that, despite being married and having three children, the candidate manages work-life balance and does not have any issues with traveling for work. The fact that the candidate changed her career path from a first-level position in another sector to a fourth-level position in the automotive sector made participants question her adaptability. The passion for the sector was highlighted in the cover letter. However, it was also observed that starting from scratch at her age might still create prejudices.

The discussion over the cover letter (**activity 3**) led to the following considerations:

- Group 1 -> The candidate looked like a rich man in the picture, but the group wondered if this was really the case. The fact that the cover letter mentioned that the candidate was a "commander" while he was doing military service was considered positively.
- Group 2 -> The candidate was considered as very much suitable for the position. According to the HR, the cover letter was written with passion and his interest for the automotive sector clearly emerged. His entrepreneurial competences were positively evaluated, while not mentioning personal hobbies and activities was seen negatively. The fact that the candidate was gay was not mentioned in the letter. The fact that the person is gay should not be considered as a problem for the company but can cause problem while working with clients and suppliers.
- Group 3 -> The picture of the candidate was considered not professional. Since she had just married, she might want a baby, and this was evaluated negatively (gender bias). The fact that she likes travelling raised her possible desire to work abroad in the future and this represented a risk for the company.
- Group 4 -> The fact that the candidate wanted to move from a sector to another raised a doubt about her ability to adapt. The fact that the candidate is a mother of three children and that was wearing a head scarf in the picture created a prejudice. Besides family issues, the group wondered if the candidate could feel comfortable in travelling due to her religious beliefs, especially for what concerns having to socialize with the opposite sex as well as to participate to working dinners with alcohol.

During **activity 4**, group 1 decided to replace the sentence about the candidate not having any travel barriers due to family issues with a more general statement mentioning that there were no travel barriers. Group 2 added some missing information but did not make any significant rephrasing. Group 3 emphasized the candidate's goals to prove her interest in staying in the company. Group 4 eliminated the sentence, "I would like to inform you that I have no travel barriers as part of the necessity of the job," as it was considered biased.

Further interesting highlights from the workshop

The association between being activists in organizations dealing with gender issues and the fact of being gay is widespread in Turkey.

Gender biases related to candidates having children were identified, as having children could imply their inability to focus on the job. The words "children" and "married" could generate both positive and negative biases.

Political-religious biases are present in Turkey, as evident from the discussion regarding a candidate wearing a headscarf.





The automotive sector is usually male-dominated in Turkey, and female employees might not receive respect from clients due to their gender.

Key takeaways from the Turkish co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 17 Key takeaways from the 1st Turkish co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Sexual orientation of the candidate	Negative bias towards gay candidates (assumed from candidates being active in gender and culture-oriented NGOs)	
Image of the candidate	Negative bias towards women wearing head scarf -> travel issues related to culture	
Family situation	Negative bias towards female candidates just married or mothers	“two children” “three children” “marriage status”
Disability of the candidate	Negative bias towards disabled candidate connected to frequent travels requested by the job	
Previous experience of the candidate	Negative bias towards previous experiences abroad of the candidates interpreted as a sign that the person could still leave the country Positive bias towards candidates having done military service	“military service” “work experiences abroad”

8.3.6 1st co-creation workshop in Estonia

Introduction

Digiotouch (DIGI) organized the first co-creation workshop on the 19th of June 2023 in Tallinn.



Figure 8 1st co-creation workshop at DIGI

Despite 25 people registered to the workshop, 20 of them participated. Participants represented the following categories:

Table 18 Categories and numbers of stakeholders at DIGI's first co-creation workshop

Categories	Number
HR officers/managers	3





AI specialists	5
Workers	9
Workers' representatives	2
Representatives of NGOs, networks, organisations fighting against discriminations	1

In terms of gender balance, 13 participants self-identified as women and 7 as men.

Plenary discussion

The open discussion was structured as a panel discussion and moderated by the facilitator. The Italian case study mentioned in the methodology was adopted along with the set of five questions proposed. Most of the participants were engaged in the discussion providing their points of view. An interesting discussion on “fairness in AI” took place among the group during which everyone agreed that fairness depends on the context in which AI is used.

The questions posed and the main points of discussion are summarized below:

1. Question 1: Based on your knowledge and experience, do we have good reasons to be optimistic or should we be more concerned with the situation in our country?
 - ➔ It was pointed out that bias exists in academia-industry collaborative projects, as industries tend to collaborate with already well-known research groups. While the training of language models on millions of data leads to optimism, bias in historical data remains a concern. Some noted that HR officers lack adequate guidelines on evaluating non-linear career transitions, leading to discrimination. In Estonia, which boasts the highest number of Unicorns per capita, most of the founders are males who demand non-stop work for competitiveness. Typical grounds for discrimination in Estonia include gender, disabilities, and hobbies. The participants agreed that human involvement in AI algorithm training is necessary, and that we shouldn't become overly dependent on AI solutions.
2. Question 2: What is your opinion and/or experience on the use of artificial intelligence systems in recruitment and Human resources management in general? Are PROs extreme tech-innovation geeks? What points would you recommend paying attention to balance the CONS?
 - ➔ LinkedIn and other AI-powered hiring platforms support recruiters in identifying suitable candidates. It was pointed out that testing and validating AI powered hiring tools with close-to-real hiring environments is necessary to understand how the tools deal with hiring decisions and adapt the technology to the national legal framework.
3. Question 3: the third question was posed to an experienced partner with a long experience in recruitment. She also provides consultancy to companies on recruitment. *What is your view on the role AI-based technology can play to favour or to hamper EDI in hiring processes in particular?*
 - ➔ AI technology, in its current state, may have a little impact on EDI in the hiring process. However, AI experts suggested that with proper trustworthiness analysis, legal framework consideration, and debiased design, AI can lead to fair recruitment outcomes. To make AI-based tech fair and inclusive, some key considerations include: 1) Fair processes supporting diversity, 2) Built-in transparency to ensure EDI from the start, and 3) Using AI tools to assess both the employer and candidate experience throughout the hiring process.
4. Question 4: how would you define it and to what extent such definition is context dependent in your view?





- The notion of fairness is context dependent, complicated to assess, and impossible to define without knowing the actual hiring context. It was noted that fairness applies both to the hiring process and the authenticity of the candidate's profile. To ensure fairness, AI must be used responsibly and regulated.
- 5. Question 5: what is your view on this? How participation of workers and social partners but also civil society organizations representing minorities can contribute to influence and oversee the use of AI in recruitment and make it fairer?
 - It is generally seen as positive to have the voice of workers, social partners, CSOs, and trade unions consulted in the design of AI systems. However, it was also pointed out that many NGOs and CSOs lack the capacity, whether financial or in terms of human resources, to influence the use of AI in recruitment and making it fair.

Group work

Four mixed groups were formed, each consisting of at least one AI expert and two workers.

Digiotouch chose the 3rd scenario from the provided methodology, as they are a technology development company. This scenario involved a tech company looking to hire a **software engineering**.

In terms of personas, race/ethnicity and gender were chosen as the grounds for discrimination. A binary gender definition was adopted to focus on the intersections of race/ethnicity and gender due to the prevalent stereotypes associated with these categories. The following personas were created:

- Persona Group 1: A white young man, born in Estonia, cisgender, married with three kids. He has over five years of experience and holds an M.Sc. in Computer Sciences. His hobbies include yoga and football coaching.
- Persona Group 2: A white young woman, born in Italy, cisgender, and single. She has around four years of experience and holds an M.Sc. in Cybersecurity. Her hobbies include cycling, running, trekking. She is also a top influencer on Instagram and collaborates part-time with tech brands.
- Persona Group 3: An Asian young woman from Malaysia, cisgender, currently engaged and without children. She is a recent graduate in computer science with no industry experience but has completed internships and freelance web development work. Her hobbies include computer games, kickboxing, swimming, and biking. She is learning Estonian (A1 level) and quickly adapts to new technologies.
- Persona Group 4: An Asian young man from Japan, cisgender, married with a child. He has approximately 3 years of experience and a bachelor's degree in computer sciences. His hobbies include gaming and travel photography. He is learning Estonian (A2 level) and is fluent in English.

The groups showed high motivation, remained focused, and all participants actively engaged in the discussions. No strong conflicts or polarizing views emerged, but the fourth activity, involving cover letter rephrasing, required more effort.

During the **first activity**, the discussion of the job offer was straightforward. The HR officer of each group went through the expected skills, competences, and the ideal profile. Spontaneously, participants discussed potential bias resulting from the job offer formulation. Key points raised were:

- Some participants perceived the requirement of "EU resident" as a potential source of bias while others argued it could be a legal requirement related to work permits.
- Inconsistencies in phrasing were noted; for example, the job description asked for a higher level of experience while seeking a bachelor's degree.
- The entry-level salary was considered mid-range, relative to Estonian tech company salaries.
- Potential bias might result from not specifying whether the position is full-time or part-time.





- Another potential source of bias could be the assumption about candidates' knowledge of technical skills acquired at a young age.

During the **second activity** the groups elaborated the cover letters. The main highlights are the followings:

- Group 1 -> The group discussed the fact that Estonian companies cannot ask about hobbies in CVs or during the interview. The mention of yoga and trekking in the candidate's cover letter could lead to bias, as it might be perceived as potentially causing work absences. The group also pointed out that yoga might be seen as a means of managing work-related stress. Mentioning that the candidate is married with three children and involved in volunteer work could give the impression that Martin might not have enough time for the job.
- Group 2 -> Discussion revolved around the candidate being a tech influencer and how this might lead recruiters to think she doesn't have enough time for the job.
- Group 3 -> Some concerns were expressed regarding the candidate's lack of work experience, having only completed internships. On a positive note, being Asian was seen as potentially biasing the perception because Asians are often associated with proficiency in tech.
- Group 4 -> the group observed that including the word "freelancer" in the cover letter could lead to bias, as it might imply that the candidate could not secure a job in a company.

In general, all groups focused on the desired skills, and the word "leader" was identified as a potential a trigger for different biases.

The discussion over the cover letter (**activity 3**) led to the following overarching considerations:

- For a tech position, including a photo of a candidate wearing a suit may introduce bias, especially if it's a woman. Such images can lead to gender bias.
- Mentioning numerous hobbies in a cover letter raises questions about whether the candidate can effectively balance their personal interests with work responsibilities.
- Using sentences like "I believe that my key expertise is" can cast doubt on the candidate's expertise. Instead, candidates should assert their key expertise directly by saying "my key expertise is."
- The expression "got laid off" in a cover letter should be considered a red flag, as it may signal a negative perception of the candidate's previous employment history.
- Overall, the detected biases are related to various aspects, including work experience (e.g., the absence of a GitHub repository), personal life (e.g., numerous hobbies), and family status (e.g., being married with three kids).

During **activity 4**, group 1 decided to add more qualifications and motivations for applying in the cover letter. Group 2 changed the sentence "I believe my expertise is" to "my expertise is". Group 3 made various changes, including replacing "internship" with "real-world/industrial experience" and replacing "freelancer" with "previous work experience." They also addressed language skills, emphasizing fluency in English and efforts to improve Estonian. Finally, Group 4 replaced "I have worked as a freelance web developer for xx years" with "I have experience as a web developer for xx years." They also rephrased a sentence for better clarity regarding administrative tasks.

Further interesting highlights from the workshop

AI specialists did not focus on specific biases and related words, while an NGO representative pointed out potential negative biases against minorities, and HR officers highlighted specific biases stemming from photos and hobby-related words.

A specific country aspect noted was that Estonia has the highest number of unicorns per capita globally. These unicorn (start-up) companies often demand employees' dedication beyond the legal 40 working hours per week. As a result, candidates with strong family ties, particularly women, may not be selected by such companies, despite having the right skills and background.





Key takeaways from the Estonian co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 19 Key takeaways from the 1st Estonian co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Family situation	Negative bias towards candidates being married with (3) children and doing volunteering -> no time for work	
Migrant origin of the candidate	Negative bias towards non-EU resident candidates -> job offer requires EU residents Positive bias towards Asian candidates -> Asians are good in tech	"Japanese is my native language and I am currently at an A2 level in Estonian"
Hobbies of the candidate	Negative bias towards candidates having "dangerous" hobbies (e.g. trekking) -> absent from work Positive bias towards candidates practicing yoga -> manage stress Negative bias towards influencers -> no time for work	
Previous work experience of the candidate	Negative bias towards candidates working as "freelancer" -> previous companies did not want to hire them	"I believe my key expertise is" "freelancer" "part-time"

8.3.7 1st co-creation workshop in Switzerland

Introduction

The Bern University of Applied Science (BFH) organized the first co-creation workshop on the 26th of June 2023 in Bern, in collaboration with the Competence Centre for Diversity & Inclusion of St. Gallen which acted as facilitators.

25 people registered and 22 participated. Participants represented the following categories:

Table 20 Categories and numbers of stakeholders at BFH's first co-creation workshop

Categories	Number
HR officers/managers	4
Representatives of HR networks	3
AI specialists	5
Workers	3
Workers' representatives	2
Representatives of NGOs, networks, organisations fighting against discriminations	5





Represented NGOs operate in the following areas: women's rights, diversity in general and LGBTQIA+ rights. In terms of gender balance, 17 participants self-identified as women, 4 as men and one person preferred not to say.

Plenary discussion

Facilitators posed to participants four out of the five questions suggested in the methodology.

The questions posed and the main points of discussion are summarized below:

1. Question 1: Based on your knowledge and experience, do we have good reasons to be optimistic or should we be more concerned with the situation in our country?
 - ➔ It was observed that Switzerland lags behind in many DEI topics, particularly concerning LGBTQIA+ issues and small companies. This is partly due to the legal framework in Switzerland, which lags behind that of the EU.
2. Question 2: What is your opinion and/or experience on the use of artificial intelligence systems in recruitment and Human resources management in general? Are PROs extreme tech-innovation geeks? What points would you recommend paying attention to balance the CONs?
 - ➔ It was observed that ChatGPT is widely used by HR professionals. Participants were skeptical about whether AI will reproduce bias and wonder if AI can reproduce empathy.
3. Question 3: what is your view on the role that AI based technology can play to favour or to hamper EDI in hiring processes in particular?
 - ➔ AI-based recruiting systems require that candidates are aware of how AI works; otherwise, it is not fair. There is a need for "augmented intelligence" rather than artificial intelligence, and humans should be involved in the decision-making process. Other tools, such as videos and voice messages, should be used in the recruitment process.
4. Question 4: How would you define it (fairness) and to what extent such definition is context dependent in your view?
 - ➔ In participants' opinion, HR processes cannot be fair. They questioned what constitutes "fairness". Apart from ensuring that everyone has the same chances of getting a job, other elements should be considered, such as who is already part of the team, and the prioritization of internal candidates. Another option could be to create different versions of the same job ad, targeting different groups, but this would require significant of efforts.

Group work

Four mixed groups were created, each consisting of five or six participants. In each group, there was at least one HR officer and one AI specialist. One group conducted the discussion in English due to language restrictions among some group members.

BFH slightly adapted the **Software Engineer** job ad provided in the methodology. The ad included both engineering (development) and more creative tasks (UX-Design). It described the tasks of the position, the required qualifications (education, experience), and skills.

As for the personas, considering the specific cultural context of Switzerland, the team adopted two dimensions: gender (male vs. female, both cis, as gender diversity remains a significant challenge in Swiss business) and country of origin/migration background. Specifically, two of the characters were born in Kosovo, reflecting the large Kosovar minority in Switzerland that still faces workplace stigmas. The profiles were as follows:

- Persona Group 1: Nadine, born in Switzerland, Swiss nationality, female cis-gender, married, 2 children. Similar experience to Fatime, but more years, albeit part-time. Bachelor's degree completed. Hobbies include yoga, school care, being team-oriented, and acting as a "mediator."





- Persona Group 2: Reto, born in Switzerland, Swiss nationality, cis-gender, no children, single. Some work experience in the field, completed a bachelor's degree, and a Master's degree was interrupted. Hobbies include traveling, volunteering as a firefighter, racing bikes, and maintaining a positive attitude.
- Persona Group 3: Bashkim, born in Mitrovica, Swiss nationality, male cis-gender, 2 children, divorced. Similar work experience to Reto, a completed bachelor's degree. Hobbies include weightlifting, camping, and being team-oriented.
- Persona Group 4: Fatime, born in Pristina, Swiss nationality, cis-gender female, married, no children. Some experience in the field, with a completed bachelor's degree. She manages a social media account about healthy baking and is sociable and human-centered.

In all groups the HR officers dominated the discussions, attempting to steer the conversation in their direction. The cover letter development activity was seen as challenging, and groups struggled to remain focused on the task.

During the **first activity**, the groups shared that the job offer was too lengthy and confusing, and that the company didn't seem to prioritize D&I or employees' work-life balance. Key discussion points included:

- Group 1: Typical male adjectives were found to be missing, while some phrases in the ad conveyed gender-specific connotations (e.g., "getting the job done" has a very masculine connotation, while "demonstrated commitment to positive customer experience" tends to have a female connotation). The group noted that many ads create the impression that candidates must be capable of everything, which might discourage women from applying if they do not feel they meet all the requirements. The job offer title was also seen as problematic due to its masculine connotation.
- Group 2: The group identified a contradiction in the advertisement, which made it unclear whether it focused on supporting or developing websites. Important details, such as salary information and company background, were missing.
- Group 3: The group found the job ad oriented toward male applicants, and the fact that it was a junior position was seen as biased.
- Group 4: The group found the job ad confusing and chaotic, without a clear differentiation between hard and soft skills. The excessive number of required skills was likely to discourage women from applying, as they tend to seek a perfect match to the profile.

During the **second activity** the groups elaborated the cover letters, resulting in these main highlights:

- Group 1: Participants discussed whether the applicant's reduced workload due to family responsibilities might have led to fewer years of work experience, putting her at a disadvantage.
- Group 3: The group observed that the Kosovar background of the candidate could be seen as both positive (indicating problem-solving skills) and negative (a stigma). The fact that he is divorced with two kids was not seen as a problem, but rather as a sign of responsibility.
- Group 4: The group mostly focused on skills in the development of the cover letter, largely ignoring aspects such as gender, origin, and marital status.

The discussion over the cover letter (**activity 3**) led to the following considerations:

- Group 1: They observed that a picture of a laughing person in the cover letter can create a positive bias. The candidate's attractiveness was seen as both positive and negative, and her appearance was deemed "non-IT typical." The fact that she is Swiss could represent a positive bias.
- Group 2: The interruption of the candidate's master's degree was viewed negatively, as it might imply future departures. A good work-life balance was seen positively, reducing the risk of burnout. The candidate's commitment to the fire department and passion for traveling had both





positive and negative implications. However, the hobby of racing bikes was perceived as a negative bias, suggesting introversion. Childlessness was considered both positive and negative, and the immediate availability could lead to negative bias.

- Group 3: The candidate was perceived as having an "alpha bro" personality, exhibiting "toxic" masculinity due to his weight training and traditional appearance. Some group members noted that these biases were connected to his origin and name, suggesting that Swiss origins might not have raised such concerns. The fact that he is divorced could be seen as a negative bias when combined with his name. The group suggested that he should have mentioned in the cover letter his desire to work remotely two days per week.
- Group 4: The group discussed the relevance of mentioning migration background in the cover letter. The candidate's attractiveness in the picture was seen as both positive and negative bias. Concerning her family situation, the group wondered if she might want kids soon (negative bias) or if she is career-oriented (positive bias).

During **activity 4**, groups discussed how to improve the cover letter. Notable discussions included:

- Group 1: Mentioning the age of the child and the skills gained from motherhood could be perceived as positive bias.
- Group 3: It highlighted that having two children was positive for a man, but if it were a woman, it might be viewed negatively.
- Group 4: The group considered that having children might indicate lower performance but could motivate employees due to increased responsibilities.

Further interesting highlights from the workshop

"Traditional Swiss-ness" emerged positively in most groups. Candidates born in other countries, such as Kosovo, were not considered fully Swiss and were required to prove their Swiss identity. Negative biases towards men from ex-Yugoslavia were also still prevalent.

Key takeaways from the Swiss co-creation workshop

It is possible to summarize the main points of discussion emerging from the discussion in the work groups in the following topics/aspects and related identified bias.

Table 21 Key takeaways from the 1st Swiss co-creation workshop

Topics/aspects discussed	Kind of bias/comments	Controversial words/sentences that fostered discussion
Job offer formulation	Negative bias towards female candidates -> female usually want to exactly match the profile	"get the job done", "demonstrated commitment to positive customer experience"
Gender or the candidate	Negative bias towards female candidates -> possible children Positive bias towards female career-oriented candidates	
Family situation	Negative bias towards candidates being married with children -> less work experience Negative bias towards divorced candidates (associated with origins) Positive bias towards male candidates having children	"divorced" "no children"
Migrant origin of the candidate	Positive bias towards Kosovar migrants -> problem solving attitude	"Kosovar origins"





	Negative bias towards Kosovar migrants -> “stigma”, toxic masculinity (associated with doing weigh training) Positive bias towards Swiss candidates	
Picture of the candidate	Positive bias towards laughing and good-looking candidates	
Hobbies the candidate	Negative bias towards candidates with bike racing as hobby -> introvert Negative bias towards candidates doing weight training -> toxic masculinity (connected to Kosovar origins) Positive bias towards candidates having travelling as hobby	“Alpha bro personality” “weight lifting”

8.4 Workshops’ core outputs: the wordlists

As explained in the methodology, the main output of the workshops is represented by wordlists that partners elaborated using the template previously made available by SVEN. In this paragraph we provide an overview of the words/sentences collected through the spreadsheets as well as an explanation of their use by WP3 tasks leaders.

8.4.1 Overview of the collected wordlists

Overall, **389 words/sentences** have been collected by partners and included in the spreadsheets. Partners were asked to indicate if each word/sentence was leading to a positive or a negative bias: it is interesting to observe how the amount of words/sentences that have been tracked as leading to positive bias (181 words/sentences) is equal to the amount of words/sentence leading to negative bias (182 words/sentences). For 26 words/sentences, partners did not specify whether they were leading to positive or negative bias, in some cases indicating that this would depend on the context.

In addition to the 389 identified words, 59 words/sentences have been reported as ‘controversial’, meaning that an agreement was not found within the group on whether they were leading/subject to a positive or negative bias, while 58 have been object of rephrasing by groups during the fourth group work activity (additional information on the rephrasing activity under paragraph 5.3 and in each partner’s report).

As explained above in paragraph 5.2, partners mainly focused on gender (all of them but one with a binary approach) and race/ethnicity as grounds of discrimination to address through the personas elaborated for the purpose of the group work. Out of the 389 words/sentences, 38 were clearly identified and reported by partners as leading to gender bias, while 48 as leading to race/ethnicity bias. Three words, instead, were classified as ‘intersectional’ leading to both gender and race/ethnicity bias. The lack of details in the dataset as far as the links/association between biased words/sentences and the specific axis of intersectional inequality will be addressed in WP3 according to the procedures described in 5.4.2.

It is important to stress that only words/sentences that partners have clearly specified as leading to gender and race/ethnicity biases and have been counted here in this summary analysis.

Figure 9 below shows the distribution of gender and race/ethnicity bias between positive and negative bias.



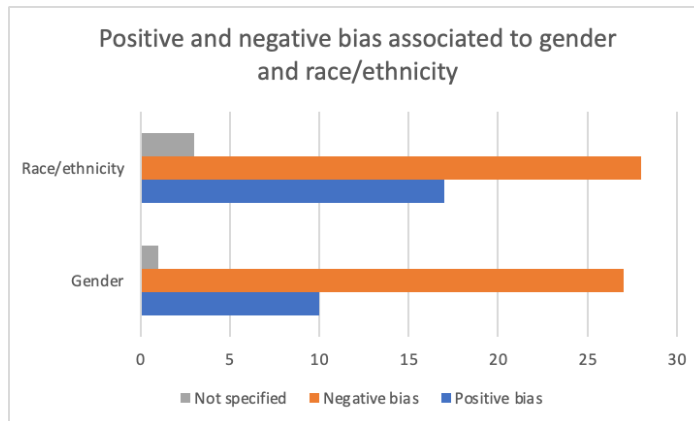


Figure 9 Positive and negative bias associated to gender and race/ethnicity

The number of negative biases is higher for both gender and race/ethnicity biases. A few words/sentences are related to other dimensions of discrimination (e.g. sexual orientation, region, age and disabilities). In particular, the table below reports the partners that have also reported in the spreadsheet words/sentences addressing other grounds of discrimination.

Table 22 Other grounds of discrimination addressed by partners in the spreadsheets

	Disabilities	Sexual orientation	Religion	Age
SVEN				X
HI		X		X
DIGI			X	
FARPLAS		X	X	X
NTNU				X
ULEID	X		X	X

As mentioned above in Chapter 4, the reporting process requested to classify words/sentences following several mostly ‘thematic’ categories, related to education and work life balance. In addition to this, an open “other categories identified by the groups” was included to allow for other intersectional axis of inequalities and related bias to emerge, so to balance the indication to stick to two main discrimination grounds (mainly gender/race and ethnicity) as requested by the emerging algorithmic modelling needs from WP3. So, the words related to Disabilities, sexual orientation religion, and age were mainly reported among the “other categories identified by the group” cluster., As far as the more ‘thematic’ categories, it might be useful to recall them as following:

- Career: work & education
- Family issues
- Work ethics
- Personal attitudes and other skills & knowledges
- Hobbies/leisure.

Figure 10 below shows the allocation of the words/sentences among the different categories detailed above.



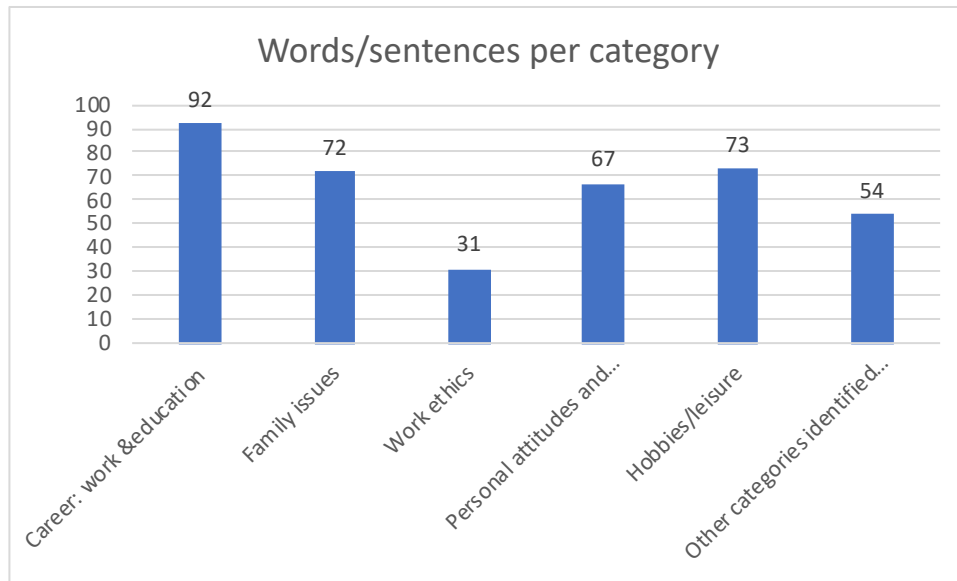


Figure 10 Words/sentences per category

As visible, the category that contains more words/sentences is “career: work & education” (92), followed by “hobbies/leisure” (73) and “family issues” (72). Figure 10 above includes both positive and negative bias, while Figure 11 below, differentiates between positive and negative biases.

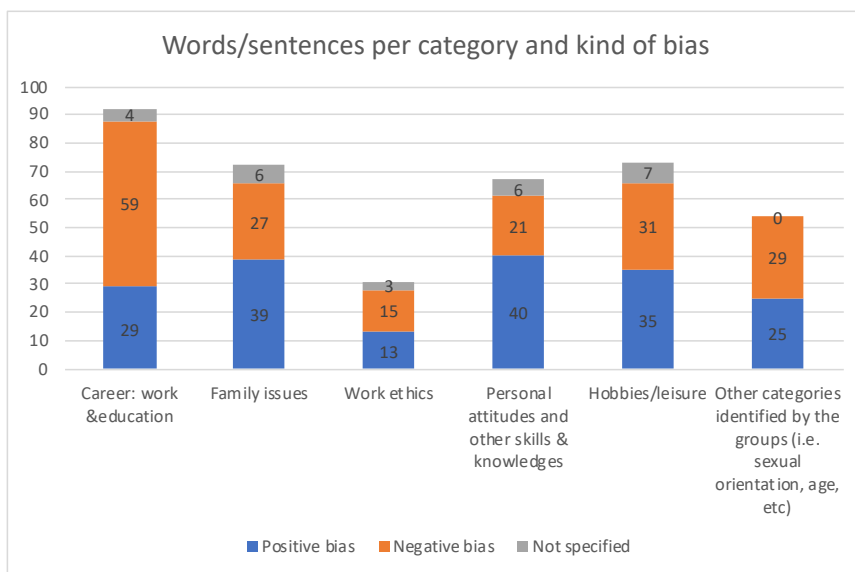


Figure 11 Words/sentences per category and kind of bias

As evident, the number of words and sentences leading to negative bias is significantly higher in the 'career: work & education' category (59 out of 92), while the opposite is observed in the 'personal attitudes and other skills & knowledges' category. Here, words and sentences leading to positive bias nearly double those leading to negative bias (40 out of 67).

The overrepresentation of words leading to negative bias in the 'career: work & education' category can be tentatively explained by the fact that many of these words came directly from the job offers themselves. The gender and/or race-ethnicity bias associated with labour market horizontal or vertical segregation in the sectors at stake may lead to the negative perception of gendered or racialized individuals. Furthermore, another relevant part of the negatively biased words was related to personas





features being over-skilled for the specific job offer, so that certain requisites, even if potentially coming with a positive bias, were perceived by participants as potentially biased in a negative sense.

The overrepresentation of words leading to positive bias in the “personal attitudes and other skills & knowledges” category, may be attributed to the tendency of potential candidates to 'overestimate' themselves by adding elements with positive traits in their CVs/profiles and cover letters, leading to their perception as positive bias by the groups.

As recalled above, it's important to note that the category 'other categories identified by the groups (i.e. sexual orientation, age, etc.)' was used by partners to report other types of bias that didn't fit into any of the other predefined categories. Specifically, under this category, biases related to gender, race, ethnicity, and other forms of discrimination, such as sexual orientation, religion, disabilities, and age, were reported when they were not associated with the predefined categories of career, family issues, work ethics, personal attitudes and other skills & knowledges, and hobbies/leisure. Additionally, partners included biases related to the physical appearance of candidates, including their clothing, as well as their expected social position.

8.4.2 Use of the wordlists in the BIAS technological development

As already mentioned in the [methodological chapter](#), the set of wordlists identified in the frame of the first co-creation workshop will support the work of the AI experts of the consortium in WP3 in relation to bias detection in static word embeddings. In particular, the objective of the next step of WP3 is to create wordlists (as in [\(Caliskan et al. 2017\)](#) and sentence templates (as in, e.g., [\(Ahn&Oh 2021\)](#)) to identify bias in word embeddings and language models, based on the interdisciplinary co-creation workshops. This helps to identify real-world bias in the language technology.

It is worth specifying that the lack of details on the association/links between bias words/sentences and the specific dimensions of intersectional inequalities in the wordlists' dataset identified during first round of workshops, will be tackled by meetings with native speakers, and the number of words will also be extended by automatic procedures (such as e.g., automatic synonym search, or search for words with similar meaning). When not a clear dimension of bias is indicated, the technological team will try them for different target words (representing the dimensions of the bias). An exploratory approach will be adopted.

In detail, the following procedure is planned, subject to adaptations due to its exploratory nature:

1. Identification of the target words (e.g., male and female words, or typical first names of specific groups for the local language or region, as done in [\(Kurpicz-Briki, 2020\)](#) for German and French). For this, existing work and native speakers will be involved to identify the respective wording, and help with language specific challenges, e.g., in German “sie” for “she” can also have other meaning such as “they”. This step is independent from the co-creation results.
2. Identification of biased words and their counterparts. These biased words will be extracted from the collected word lists of the co-creation workshops directly, or indirectly by using synonyms and similar words. In this process, the original workshop notes as well as native speakers will be included as needed, e.g., with focus groups.
3. The different wordlists and sentence templates will be fed to existing methods to measure bias (e.g., WEAT [\(Caliskan et al. 2017\)](#)), but for the word embeddings and language models in the local languages. On one side, this will give insights whether this real-world bias can be confirmed in the word embeddings and language models, and on the other side this enables the adaptation of the methods to measure bias to the specific challenges of the local languages and cultural aspects. Due to these challenges, it is expected that not all created word lists and sentence templates will show bias in the language models and word embeddings. The ones that allow to measure such a bias, in the long-





term, will help to test the language specific bias mitigation methods for word embeddings and languages models to be developed later in the project.





9. Methodology of the second co-creation workshop

9.1 Workshop's agenda and target

The second workshop had two main purposes:

- Discussing on fairness in the first phase of the recruitment processes, in particular in the screening process: identification of fairness principles and features of a fair recruitment process. Prioritizing candidates' features and required qualification/skills for a job offer.
- Identifying desirable requirements and functionalities of a Debiaser tool and a CBR based decision making support system and related risks.

The workshop involved ideally 24 stakeholders, and in this round of co-creation HR officers and specialists were prioritized. For the special focus on the notion of fairness, two additional types of profiles were added, namely philosophers and legal experts (on Human Rights and/or Labour Law). Contribution from philosophers was deemed useful and important to add critical perspectives on fairness definitions. Legal experts' points of view could add value both in terms of the specific implications of unfair procedures in selection/recruitment, and the identification of requirements and the evaluation of AI systems and the Debiaser in particular. A balanced group composition was presented as per Table 23 below:

Table 23 Categories and ideal numbers of stakeholders involved in the 2nd co-creation workshop

Type	Ideal number
HR officers and networks, associations of HR specialists preferably already active on gender/diversity & inclusion issues	10–12 people
Representatives of civil society organisations (e.g. associations, NGOs), networks, organisations fighting against discriminations (in particular, but not exclusively related to gender and race)	2–4 people
Legal experts in human rights and/or labour law	2 people
Philosophers	2 people
Workers and workers' representatives (e.g. trade unions)	2–4 people
AI specialists	4 people

It was suggested that particularly interested and motivated participants from the previous workshop were also invited to the second one. Still, given the different share of stakeholders per typology, partners had to focus on engaging more HR officers as well as participants with the two new types of expertise sought for this co-creation session.

The event lasted approximately four hours, and was structured shown below in Table 24:

Table 24 Structure of the second workshop

Programme	Methodology	Timeframe
Participants' welcoming/introduction and BIAS presentation	Plenary	30 minutes
How does a fair HR recruitment process look like?	Discussion in two groups and plenary discussion	70 minutes
Which requirements for AI tools in recruiting?	Interactive/hands-on work in four groups and plenary discussion	80 minutes
Lunch/aperitif/dinner		60 minutes





Also in this case, workshops were preferably conducted in presence, as networking was identified as one of the main incentives for participants to join.

No dedicated coffee break was foreseen during the workshop, therefore it was suggested that partners prepare a corner with coffee, water and snacks available throughout the whole duration of the workshop.

9.2 Introduction & BIAS presentation

The first 30 minutes of the workshop aimed at welcoming participants, allowing introductions as well as presentation of the BIAS project, considering that part of the stakeholders was different from the ones participating in the first co-creation workshop.

Within the introductory session, information were shared with participants both about the results of the previous co-creation workshop as well as about BIAS' next steps and how the results of the workshop will be used. In particular, on BIAS next steps, it was suggested to highlight the following:

- a) A third **international co-creation workshop** that will be organized on **December 7th 2023**, in Venice with project partners and up to 3 stakeholders per partner (apart from LOBA and CrowdHelix) validating results of the two previous workshops and advancing with the Debiaser requirements and functionalities co-design.
- b) Results from the full co-creation cycle will be analysed in a report that will be public. Even before the release of the final deliverable, BIAS computer scientists will rely on co-creation results when programming the Debiaser's AI models.
- c) In addition, training courses and raising awareness events will be organized by BIAS starting from October 2023.
- d) In order to be kept updated with all BIAS project development and the above-mentioned events, participants are invited to subscribe to the national BIAS Lab.

9.3 Discussion in two groups: how does a fair HR recruitment process look like?

The first workshop activity consisted in a discussion in which participants were divided in two groups to facilitate dialogue and allow everyone to contribute to the discussion.

Two balanced groups in terms of stakeholder categories were created. It was suggested that each group should be ideally composed of:

- 5–6 HR officers
- 2 AI specialists
- 1-2 workers' representatives
- 1 philosopher
- 1 legal expert
- 1–2 representatives of civil society organizations

The groups had 45 minutes to discuss over the following points/questions, and 25 minutes overall to report and exchange in a plenary session.

The overall topic of the discussion was "fairness in HR recruitment processes".





The two groups were moderated by two facilitators, who introduced the topic of the discussion through a set of slides using the contents below.

A first slide provided the following information followed by a first question:

Slides contents:

There isn't consensus on a single "human definition" of fairness, but many overlapping and conflicting definitions exist which are often "sector sensitive"¹¹.

In HR recruitment/selection, fairness problems/issues revolve around the following main points¹²:

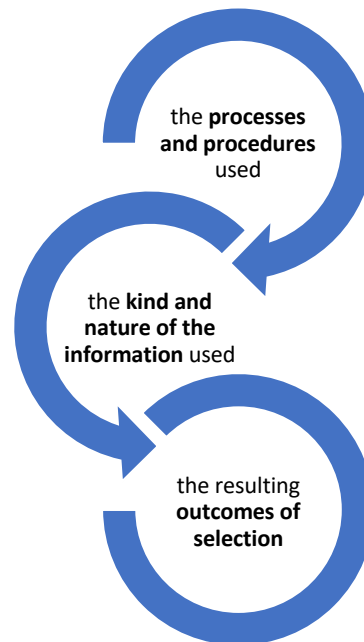


Figure 12 Main points concerning fairness in HR recruitment/selection

Regarding the first block "**the processes and procedures used**", and as pointed out in D2.1, relevant literature¹³ has identified the following as the most common principles of procedural fairness:

1. Objectivity -> subjective decision making from the recruiter/employer is minimized. An objective selection procedure is featured by the use of quantitative methods (e.g. test scores), more formalised decision-rules and criteria for selection.
2. Consistency -> all applicants are treated the same, receiving the same interview procedure, the same tests, etc. Examples of violations of this principle is when males and females are asked different questions during interviews.
3. Non-manipulation -> each applicant should be examined against a common set of criteria and standards (non relevance of factors such as political ties)
4. Professionalism -> the selection process is put in place by professionals
5. Job relatedness -> the recruitment and selection procedures should only assess the personal characteristics that are necessary for the job and can predict the skills and capabilities of the job applicant.

¹¹ See the paper "An introduction of the fairness notion for BIAS-project people" prepared by Pinar Pinar Øzturk from NTNU, available at the following [link](#)

¹² Arvey R.D., Renz G. L, "Fairness in the selection of employees, Journal of Business Ethics. 11 (5-6):331-340 (1992)

¹³ (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) (Arvey, Renz, 1992),





6. Multiperspective -> The review of applicant information and selection decision is made by several individuals who represent different perspectives and constituencies
7. No discrimination -> Candidates are not discriminated against because of their gender, sexual orientation, race and ethnicity, religious belief, different abilities, etc.

1st activity: considering principles 1, 2, 3, 5, 6, 7 (since professionalism should be already universally recognized as fundamental), which one would you consider more important and which one less important in the candidates' screening phase? Facilitators proposed participants to answer using an online polling system (i.e. [Mentimeter](#)) by framing the QR code included in the slide.

After the vote in each group, a brief discussion follows if any of the participants want to comment on the result of the poll (5 minutes).

2nd activity: as principles 1 and 2 are the most relevant for the BIAS project when designing the CBR system, facilitators asked participants in groups how they concretely implement those in the applications' screening phase (this question was addressing HR officers in particular). Which procedures, measures, tools and data/info are used to ensure objectivity and consistency? (10 minutes).

3rd activity: facilitators engaged the two groups in an activity simulating a CV screening process. A scenario with a job offer and a company profile was provided to each group with three personas, or fictitious candidates profiles. The goal of this exercise was to elicit the reasoning behind a recruitment process for what concerns the screening phase.

The groups received the following materials that can be found in [Annex 8](#):

- a job offer with company profile (a different job offer per group)
- 3 fictitious candidates for the position (three different fictitious candidates per group, 3 for one group 3 for the other).

Partners decided whether to translate in local language the material, or to adapt it to better align to their national context. In the case they opted for adapting it (and choosing different job offers/company profiles/candidates), the following aspects had to be taken into account:

- Candidates' profiles: ensure enough diversity among profiles, that the key intersectional categories for the project are reflected (gender/race/gender identity-sexual orientation) and that there is not an obvious "winner" of the competition for the post (or vice versa), so to trigger an interesting discussion. Differently from the 1st co-creation workshops, no other conditions/criteria applied to the choice of the personas.
- Companies' profiles: internal recruitment practices and or HR Management policies could vary, EDI (Equality, Diversity and Inclusion) policies could be there or not, but it was suggested to have one company profile with these policies (or similar) in place and another company profile without such policies.

At least four paper copies of the material were printed and provided to each group.

The HR officers of the groups went through the material provided for around 5 minutes and then the facilitators triggered a discussion posing the following questions:

- Reflect on the provided information/variables both regarding the candidates and the company profile: which ones are important to consider in a first screening of received applications? Can you agree on an order of importance?
- Reflect on elimination criteria: is there one candidate among the 3 that you would eliminate for sure? If yes, which one? Why and how did you get to this decision? Which is the reasoning behind?





- Reflect on the selection criteria: if you have to choose a candidate to interview among the 3, which one would you? Why and how did you get to this decision? Which is the reasoning behind?
- In case it did not emerge during the conversation, explore how much important the company's rules and policies were in orienting the decision on the candidates to eliminate and to invite to an interview.

During both this activity and the previous one, the main target were HR officers, while the other members of the group (philosophers, AI specialists, NGOs/trade union representative, legal experts) simply observed, taking notes and reporting on their reflections during the plenary. Facilitators highlighted the importance of tapping into different perspective on the same issues from participants with different background. They were instructed to consider if and how in their opinion, the outcomes of the discussions in the groups affected/had an impact on workers' rights and if and how an AI-based technology could take into account the different needs/reasonings behind a screening process.

A rapporteur in each group supported the facilitator and took notes on the results of the discussion using the template available at [Annex 9](#).

A 25-minute plenary session followed organized as following:

- each group briefly presenting the main results of the discussion (5 minutes x 2)
- participants different from HR officers from the two groups taking the floor with their short feedback on the discussion (15 minutes)

9.4 Interactive/hands on work: which requirements for AI tools in recruiting?

Participants were split in four groups (six people per group), each group ideally having three HR officers, one AI specialist, and two people from civil society organizations, one workers' representatives, philosophers and legal experts.

The group work explored ideal requirements of the Debiaser, in its different language bias detection and mitigation component and the decision-making support system drawing on "Case Based Reasoning".

The work was introduced by a brief explanation by facilitators about the two different models using a summarized/shortened version of the information already made available by BHF and NTNU.

The presentation of the two tools should have lasted 8–10 minutes. After the introduction, participants were split in the four groups as mentioned above. It was suggested to print four copies of the presentations and distribute them to the groups.

The next three sections explain:

- The "future-state journey map" technique that inspired this exercise.
- How the "future-state journey map" steps were "reinterpreted"/adapted for the purposes of the exercise.
- How the exercise was actually implemented/developed.

1. The "future-state journey map" technique

The group work simulated a "recruiter's journey" towards the selection of a candidate for a given job offer. The technique which inspired this part of the workshop is the "future-state journey map"¹⁴, primarily used by companies to understand and improve their existing customers' experiences. The aim of such technique is to articulate a "vision" rather to record an existing journey. The main steps of developing a "future state journey map" are the following:

- a. Forming a mapping stakeholders' group

¹⁴ <https://www.mindtools.com/aiwijpy/designing-future-state-customer-journeys>





- b. Mapping the current-state customer experience
 - c. Defining business goals and target customers
 - d. Generating new ideas
 - e. Mapping the future-state customer experience
 - f. Validating the map
 - g. Putting the map into work
2. **Adapting the “future-state journey map” technique to the co-creation exercise**
- a. Forming a mapping stakeholder group

As already mentioned, first of all **balanced groups of stakeholders** (3 HR officers, 1 AI specialist and two people among civil society organization representatives, workers’ representatives, philosophers and legal experts) were created by facilitators (it is recommended to pre-prepare the groups in advance in order to save time).

In each group a facilitator/rapporteur was present to guide to the groups, manage time and take notes.

To develop the exercise, a common persona character and scenario were defined. This time, the persona was a HR Officer/Manager, rather than a candidate.

Table 25 Scenario and persona of the 2nd co-creation workshop

Persona	Scenario
HR officer of a company working in the retail sector and having 10.000 employees	The company is looking for an administrative assistant. It receives around 1000 applications for the position. The company currently has an ATS that supports administrative tasks: it manages candidates, schedules job interviews and send emails.

- b. Mapping the current-state recruiter experience

The group work focused on the screening phase of the recruitment process. In particular, the groups brainstormed on the following basic steps of the screening phase.¹⁵

1. Ticking off the basic or must-have requirements (included in the offer).
2. Scanning for preferred or good-to-have qualifications.
3. Matching the holistic picture of the candidate to the role.

A poster template with the three above-mentioned steps was elaborated to be used during the group work (see the version in English below). Partners had to translate and print it in A1 format. The different links to the translated versions can be found in section 3) “group work development”.

¹⁵ <https://www.spiceworks.com/hr/recruitment-onboarding/articles/what-is-candidate-screening-and-selection/>



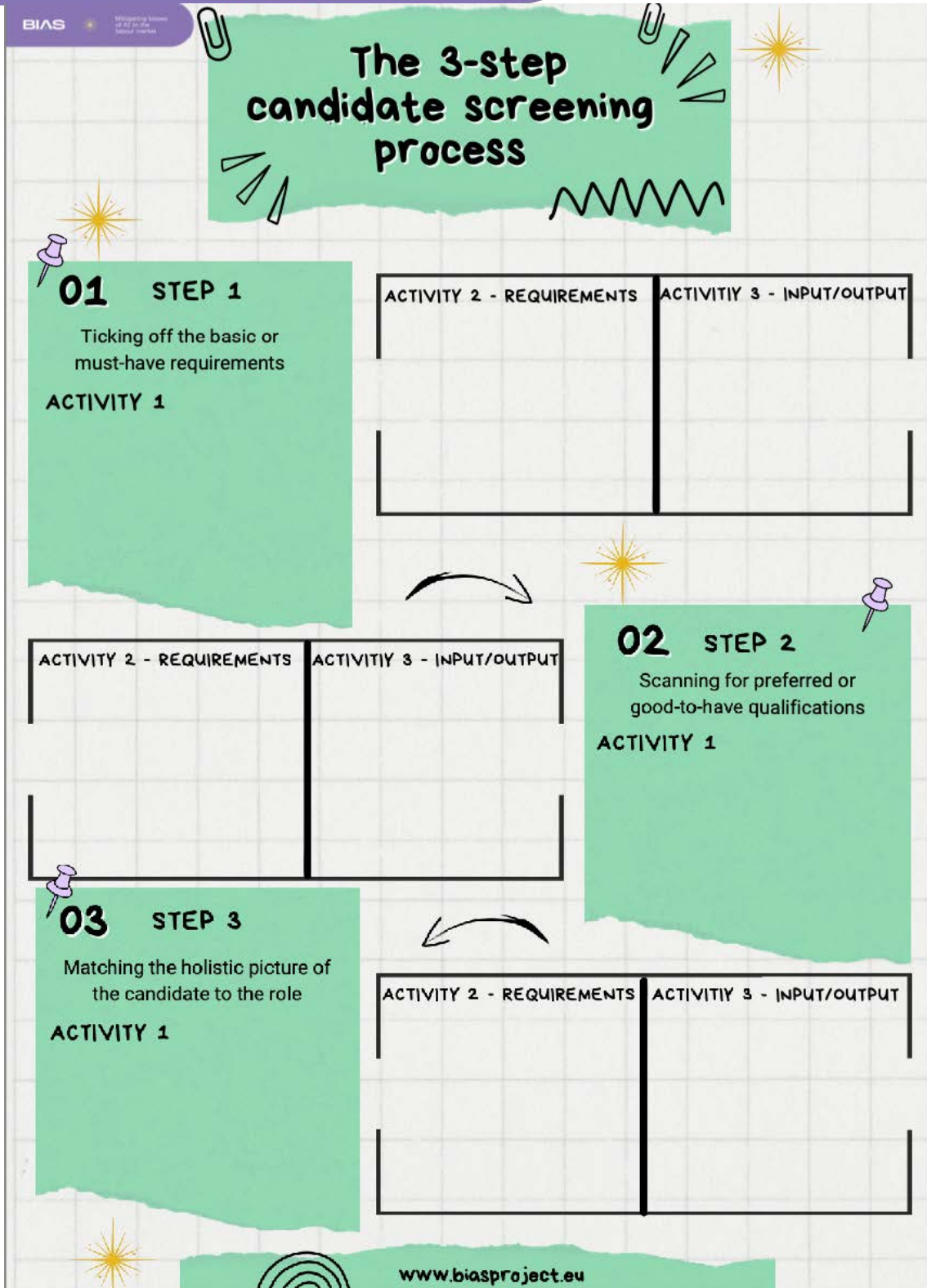


Figure 13 Poster template for the 2nd group work exercise in the 2nd co-creation workshop





c. Defining the goals

The goal of this activity was to map the future state recruiter experience in the applications' screening phase and identify requirements of an AI-based Debiaser tool. The groups identified how an AI tool could support/improve the work of a recruiter during the phase of screening CVs and cover letters process. The exercise aimed at identifying ideal requirements of an AI tool based on the technologies developed in the BIAS project in order to support the recruiters' work. Facilitators highlighted how the Debiaser tool was to be imagined in its different components, both language bias detection and mitigation and the "Case Based Reasoning" decision making support system. They were also advised to stress that the systems to be envisaged could process written information/texts only (no video-audio input feeds into the system).

This represented the core activity of the group work. Starting from the **current-state of a recruiter experience** (point 2), participants were asked to reflect on the screening process, in particular on the screening phase focusing on desired/imagined requirements of such technologies.

In particular, following a system engineering approach, both functional and non-functional requirements could emerge, where functional requirements describe what a platform/tool is supposed to do, and non-functional describe how it is supposed to work. It was suggested that facilitators would clarify the difference between the two as following:

- functional requirements are usually in the form of "system shall do" requirement, for example a platform should send an email to every user that performs the registration, non-functional requirements, instead, are in the form of "system shall be" requirement, a non-functional requirement could dictate that the system is highly responsive and such email must be sent in under two seconds¹⁶. In the specific case of the BIAS technological tools, an example of functional requirement could be: the tool should screen all the received CVs and end up with a number of CVs of candidates to be interviewed and indicate any sensitive information/expression in the CVs that are at risk of causing bias.
- Connected non-functional requirements could be that the tool should have sufficient processing memory and system reliability to receive at least 1000 CVs as input, or the tool should be integrated in ATS (Applicant Tracking Systems) that are already in.

Participants were asked to identify any type of requirements and to reflect on the expected inputs and outputs of the identified technologies.

d. Validating the map

The groups validated the map during the plenary session in which the results of the group work are shared with the other groups.

e. Putting the map into work

The findings of the group work were collected in an overall report by Smart Venice and shared with the technological partners of the consortium.

3. Group work development

The five steps identified above were developed as follow.

After splitting the groups in four, the facilitators distributed the following materials to the groups:

¹⁶ Definitions of functional and non-functional requirements taken from deliverable 1.4 "Models, Methodologies, Scenarios & Requirements – Final" of the EMPATIA Horizon 2020 Project





- The poster with the three steps of the screening process (to be printed in A1 format) which was made available for each partner.
- The table for reporting activity 4 (to be printed in A3 format) ([Annex 10](#))
 - The table below with the description of the activities (to be printed in A4 format)
 - Slides introducing the Debiaser and the CBR model made available by WP3 experts ([Annex 11](#))
 - The template for the rapporteur ([Annex 12](#))
 - Sticky notes of at least 3 different colours

The facilitator presented the scenario and the fictitious character the group would work on:

Table 26 Scenario and persona of the second workshop

Persona	Scenario
HR officer of a company working in the retail sector and having 10.000 employees	The company is looking for an administrative assistant. It receives around 1000 applications for the position. The company currently has an ATS that supports administrative tasks: it manages candidates, schedules job interviews and send emails.

Facilitators had to clarify to the groups that for this exercise, differently from the previous one, no detailed scenarios and personas were provided given that the focus is on the requirements of the technological solution.

The group work last overall one hour and was structured in the following activities that are explained by the facilitators beforehand.

It was recommended to provide clear guidelines to participants before starting the exercise, especially on how to use and report notes in the poster provided. In particular, it should have been explained that the poster is structured in the 3 screening steps and for each step notes (either using sticky notes or directly writing on the poster) related to the first activities detailed below should be added.

Table 27 Structure and description of group work in the 2nd co-creation workshop

Activity	Description
First activity – brainstorming on the screening phase (10 min)	<p>The group observes the three steps of the screening phase and brainstorm on the following questions:</p> <ul style="list-style-type: none"> - How could an innovative technology based on Natural Language Processing (NLP) and Case Based Reasoning (CBR) support the three steps of the screening phase? - Which needs would the technology address? <p>The group uses sticky notes of the same colour and paste them on the poster (under activity 1) or directly writes notes in the poster indicating how the technology would support the phase.</p> <p>The rapporteur reports on the template provided making sure to also highlight any different positions/ideas of participants.</p>
Second activity – identifying requirements (20 min)	<p>Once the group has identified how an AI technology would support the screening phase (activity 1), the following step would be to identify requirements of the technology (as described above).</p> <p>In particular, per each screening step identified in the poster, the group should try to answer the following questions:</p> <ul style="list-style-type: none"> - What should the tool do? (e.g., the tool should screen all the CVs received and end up with a number of CVs of people to be interviewed and





	<p>indicate any sensitive information/expression in the CVs that are at risk of causing bias)</p> <ul style="list-style-type: none"> - How should the tool be, in order to do it? (e.g., the tool should be wide enough in terms of processing memory to receive at least 1000 CVs, or the tool should be integrated in the ATS) <p>When discussing on the different requirements the group should also try to identify any specific risks associated to workers' rights or to the technology. In particular, the AI specialist of the group will report the feasibility from an AI perspective of the identified requirements, philosophers and legal experts will highlight risks from an ethical and legal point of view.</p> <p>The group participants either use sticky notes (a different color from activity 1) for the requirements and attach them on the poster or directly write in the poster the notes.</p> <p>The rapporteur fills the report indicating the different requirements identified differentiating per each screening phase as well as any diverging positions among participants.</p>
<p>Third activity – identifying inputs and output of the tool (10 minutes)</p>	<p>The third part of the group work is aimed at eliciting, per each identified functional requirement, which are the needed inputs and expected outputs of the system, that have most likely already been identified in the previous rounds of discussions.</p> <p>For instance, in the example provided, the needed inputs are the CVs and the job offer, while the expected outputs are:</p> <ul style="list-style-type: none"> - the identified/selected CVs with the matching % between CVs' skills and job offer's skills - sensitive information/expressions and explanations <p>The inputs and outputs identified need to be reported in the poster (using sticky notes of a different color or writing directly in the poster) and in the group template by the rapporteur.</p>
<p>Fourth activity – identifying conditions for fairness/trustworthiness and evaluation (10 minutes)</p>	<p>The group brainstorms on the following two points:</p> <ul style="list-style-type: none"> - conditions/features that an AI system should have in order to be considered fair and trustworthy. - how fairness and trustworthiness of an AI system can be evaluated/measured both qualitatively and quantitatively. <p>While participants will the table (Annex 10), rapporteurs take notes on the development of the exercise pointing out at different positions emerged among participants.</p>
<p>Walking plenary (30 minutes)</p>	<p>Each group shares the results of the group work focusing on the identified requirements, needed inputs and expected outputs, conditions for fairness/trustworthiness and evaluation methods.</p>

As **careful time management during the group-work** was crucial to complete all the activities, different possible ways were proposed to ensure the groups complete the session:

- Time to be managed centrally and someone from the hosting partner's staff signals the different slots on a slide accompanied by a (gentle) sound when the time for each slot has expired.
- Rapporteurs to take this role.
- The group assigning this role to one of its members.





9.5 Reporting process

The results of the workshop had to be summarized in reports, drafted by partners' teams: as already mentioned above, rapporteurs had to take notes during the two different activities (the discussion in two groups and the group work) using specific templates made available **and in national language**. A detailed final report in English, incorporating the results of all group work, had to be completed immediately after the workshop ([Annex 13](#))¹⁷.

Delivery of the reports, both the ones in national language and the one in English was required as soon as possible after the workshops took place to minimize risks of misinterpretation, considering that conversations weren't recorded. It was recommended that rapporteurs, while taking notes during the exercise, indicate whether there was agreement or disagreement in various phases and provide the different arguments.

The overall report of the second co-creation workshop, in English, had to be sent to Smart Venice by the 20th of September 2023.

¹⁷ The structure of the overall report reflects the one of the two group reports. In case for the activity on fairness in recruitment processes, partners had used different job offers, company and candidates profiles than the proposed ones, recommendation was given to describe them in the report, translate in English and provide them as annexes to their overall report





10. Implementation of the second co-creation workshops: results

10.1 Overview of the conducted workshops

In September 2023, 7 partners (SVEN, NTNU, ULEI, HI, DIGIO, BFH and FARPLAS) conducted the second round of co-creation workshops. Overall, 163 people registered, while 131 participated. The seven workshops gathered people with diverse characteristics in terms of gender, expertise, and occupational backgrounds.

Participants and their Roles

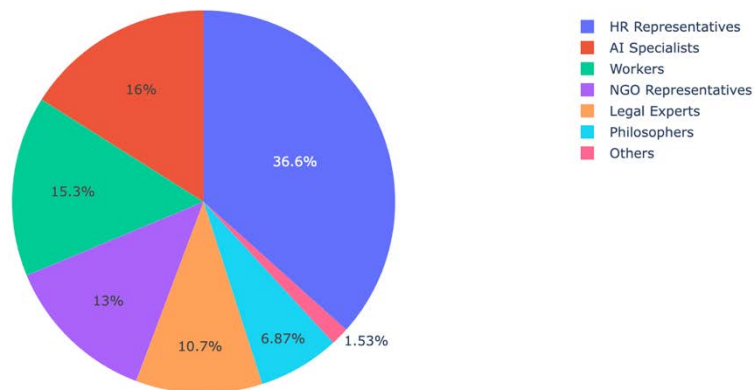


Figure 14 Distribution of fields of expertise among participants to the 2nd co-creation workshops

Diverse disciplines and organization types were represented in the group. Many participants were HR officers, followed by a significant number of AI specialists, legal experts, philosophers, and representatives of civil society organizations (CSOs). There were a few workers and workers' representatives, some of whom were also playing other roles in their respective organizations. This reflection of cross-disciplinary and overlapping expertise was quite common.

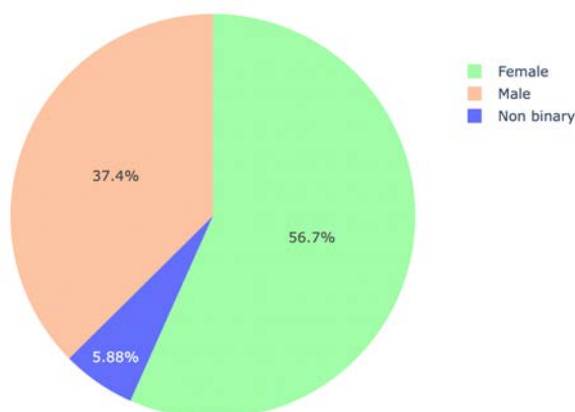


Figure 15 Gender distribution in the 2nd co-creation workshops

Regarding gender distribution, not all partners provided specific details on this aspect. Out of the seven partners, only four expressed the gender ratio of the participants. If we consider these data as an approximation of the actual gender distribution, it suggests that there were more female participants across the workshops. Among the partners that specified the gender ratio, there were no non-binary participants.





The average number of registered participants at every workshop was 23, while an average of 19 actually attended: the absence rate corresponded to 8%. Absences or early departures in some workshops were mainly due to reasons such as illness, other duties, and personal issues. In general, the workshops were inclusive and ensured diversity in their participant composition, taking into account various factors such as gender balance, professional expertise, and special needs. However, it was observed that absenteeism and early exits of participants were mostly due to personal reasons, indicating that more convenient timing or health support might be needed in future workshops.

Two workshops faced special circumstances: one accommodated a participant with a hearing impairment and ADHD (Attention Deficit Hyperactivity Disorder) by creating smaller groups and conducting separate sessions, while another had a mismatch in participant numbers between sessions due to an unexpected departure. The incomplete attendance was due to a variety of reasons from sickness to schedule conflicts.

All seven partners filled in the overall report with the results of the workshop. All of them reported the workshops being successful and meeting the expected results both in terms of the initial group discussions and the interactive exercises. In analyzing the outcomes of this workshop cycle, the results and data were consolidated across all partner entities. The nature of the topic discussed did not necessitate differentiation based on specific countries, as most of the workshop outcomes did not involve country-specific details. While exploring potential differences in mentalities among countries could be intriguing, such considerations cannot be evaluated based on the reports drafted during these workshops. It was highlighted that the scope of the activities did not include a deliberative component or a sampling component. Therefore, the considerations extracted from the discussions are not representative of a specific country. For these reasons, a country-specific analysis was considered beyond the scope of the co-creation exercise. Unlike the prior co-creation workshop, this session did not incorporate any language-related exercises, consequently, all the reports were collectively analyzed without summarizing their results in dedicated paragraphs as it was the case in D2.2. When dividing participants into sub groups, language preference was considered as a split factor by NTNU (which had a Norwegian-speaking group and an English-speaking group) and by BFH (which had a German-speaking group and an English-speaking group), but the report was always filled in English.

10.2 Discussion into groups: what does a fair HR recruitment process look like?

10.2.1 Activity one: ranking principles – Results

As the first activity of the conducted workshops, participants were asked to rank the most common principles of procedural fairness, from the most important to the least. The principles in question are: 'Objectivity', 'Consistency', 'Non-manipulation', 'Professionalism', 'Multiperspective', 'Non-discrimination'.

The principles that were considered most crucial for fairness in recruiting were 'Non-Discrimination' and 'Objectivity'. Participants emphasized the importance of tackling bias within AI algorithms used for recruiting practices, as these systems can perpetuate discrimination based on past data. Focus on job-related factors and minimization of subjective evaluations were also essential parts of ensuring fairness. Thus, 'Job-Relatedness' was also deemed crucial in evaluating personal characteristics in alignment with job requirements, while 'Objectivity' emerged as a principle fostering the minimization of biased decision-making.

Conversely, the principles that were deemed the least important were 'Consistency', 'Multiperspective', and 'Non-Manipulation'. Although these principles received fewer votes, they were still considered critical to fair and unbiased recruitment. 'Consistency' stood for treating all candidates similarly in interview procedures and tests, crucial in mitigating potential unfairness from differing standards applied to different candidate groups. 'Multiperspective' valued the input of diverse viewpoints in recruitment,





ensuring a more comprehensive evaluation of candidates. 'Non-Manipulation', although deemed less important, was considered significant in maintaining the integrity of the recruiting process.

Table 28 The results of ranking the fairness principles

	Non Discrimination	Objectivity	Job Relatedness	Non-Manipulation	Multiple Perspectives	Consistency
BFH	1	2	3	4	5	6
SVEN	1	2	3	6	5	4
HI	1	3	6	5	2	4
ULEID	1	2	4	6	5	3
DIGI TOUCH	3	2	1	4	5	6
NTNU	1	2	4	5	6	3
FARPLAS	2	1	4	6	5	3

Non Discrimination	1.43	1°
Objectivity	2.0	2°
Job Relatedness	3.57	3°
Consistency	4.14	4°
Multiple Perspectives	4.71	5°
Non-Manipulation	5.14	6°

The importance of empathy emerged in the discussions, countering the emotionless aspect of AI, which could potentially perpetuate biases. The discussion also highlighted varied perspectives on 'Job-Relatedness', especially in HI, NTNU and ULEID workshops. Some participants felt that job-related factors should not be the deciding elements in recruitment as they could be biased, emphasizing the importance of evaluating each recruitment scenario individually and in its complexity instead. Unconscious prejudices and potential effects on hiring decisions were important elements to consider, as were discussions on underlying issues like discrimination by association, or tendency to prefer others who share similar backgrounds.

NTNU groups also specified that principles' prioritization can differ between the public and private sectors. In the public sector, recruiters tend to emphasize objectivity and consistency, while in the private sector recruiters lean more towards job relevance. Participants from the public sector mentioned operating according to positive-discrimination, particularly when considering candidates' disabilities during recruitment.

In conclusion, most participants viewed the principles not as individual entities but elements of a holistic approach to ensuring fairness in recruitment. While differences emerged in terms of prioritization, the overarching consensus leaned towards fostering a recruitment process that nurtures non-discrimination, objectivity, and respect for each candidate's unique strengths and capabilities. The figure below reports the total score for every principle: the higher the score, the more frequently the principle was put in first positions. The scoring also takes into account penalties for principles that were identified as least important:





Figure 16 Graph representing the scores of every fairness principle, considering votes and penalties



Figure 17 Scores of every fairness principle, considering votes and penalties

10.2.2 Activity two: Objectivity and Consistency - Findings

In the second activity of the workshops, participants were asked to reflect deeper on the principles of Objectivity and Consistency, with particular focus on the procedures, measures and tools to implement them (both existing and potential). The seven workshops discussed the importance of objectivity in recruiting systems, and highlighted several key elements to ensure this. Using a methodology that eliminates personal bias was found to be crucial, including the use of standardized criteria, scoring systems, and structured interviews. Removing personal information, such as photos, marital status, and gender, from CV's was considered a potential way to improve objectivity. Flexibility and multiple perspectives were encouraged, including the potential of setting up panels composed of HR staff, managers, and employees when evaluating applicants. To prevent the skewing of processes by personal recommendations, equal opportunity can be promoted through public job postings with clear requirements. The application of technology such as blind recruitment and the use of software for scoring was considered beneficial, and there was unanimous agreement that decisions should be based on qualifications rather than personal characteristics.

Ensuring consistency in recruitment processes was a shared concern throughout the workshops. Maintaining a consistent screening team and providing them with diversity and inclusion training was recognized as fundamental. Standardizing measures like structured interviews, application forms, and criteria for all candidates were suggested for consistency. Additionally, technology, metrics, semi-structured interviews, and feedback from job applicants were considered helpful in maintaining consistency. The use of behavioral models in interviews and the inclusion of external and neutral personnel such as auditors were suggested to meet these standards. It was also recommended that all personal data, including photographs and marital status, should be consistently removed from all applications.

The workshop discussions recognized the possible fallacies of objectivity and consistency, indicating skepticism regarding the feasibility and effectiveness of implementing strict objectivity and consistency in recruiting systems. Still, it was argued that attempts should be made to reduce bias as much as possible. The participants discussed how diversity among candidates could sometimes conflict with the pursuit of consistency. However, the enrichment such diversity brings to the team should be noted. The idea that ensuring a truly objective hiring process could result in a homogeneous and monotonous workforce was also put forward. In general, from the workshops a common thread emerged highlighting the need for a balance in emphasizing qualifications while ensuring varied perspectives. Despite all efforts, the success





rate of finding the right candidate was deemed to be relatively low (around 50.2%), signifying that the recruitment process is still far from perfect. Participants also stressed the danger of manipulation in interviews and suggested measures like intensive testing to reveal candidates' genuine personalities. Finally, the participants indicated skepticism regarding the feasibility and effectiveness of implementing strict objectivity and consistency in recruiting systems.

During the third activity of the workshops, participants were asked to simulate an hiring process, and, based on the specifics of the job offer and the candidates' profiles, to identify the relevant criteria, and differentiate between elimination factors, and the selection criteria.

Relevant Criteria

Participants of the workshop emphasized the importance of various factors in the initial screening of candidates for a job position. The following is a comprehensive list of the most relevant criteria identified by the participants during their profile analysis

For the Assistant Store Manager position:

- ➔ Meeting the basic requirements specified in the job offer
- ➔ Relevant professional experience
- ➔ Duration of previous experiences
- ➔ Motivation and cultural fit
- ➔ Communication skills
- ➔ Education compensating for experience
- ➔ Knowledge of languages
- ➔ Availability on Saturdays and Sundays
- ➔ Teamwork and social skills
- ➔ Legal regulations compliance
- ➔ DEI policy compliance (Diversity, Equity and Inclusion)
- ➔ Job description clarity
- ➔ The home-work distance
- ➔ Having blind CV for initial screening
- ➔ Additional relevant information included having worked in a sector similar to the job's sector, other soft skills, not being overqualified, time flexibility, hobbies, additional licenses and training, educational background, dynamism.

For the Logistic Officer position:

- ➔ Meeting the basic requirements specified on the job offer
- ➔ Living close to the workplace
- ➔ Professional Experience
- ➔ Educational background
- ➔ Good health
- ➔ Night Shift availability
- ➔ Flexibility
- ➔ Motivation and cultural fit
- ➔ Job description clarity
- ➔ Communication skills
- ➔ Previous experience within the company
- ➔ Having blind CV for initial screening
- ➔ Additional relevant information included having a diploma, not being overqualified, and dynamism.





Participants agreed that global company experience, education, and industry experience are crucial. They also deemed skills, qualifications, and alignment with the company's requirements as top priorities, suggesting AI tools could aid in filtering out candidates based on these factors.

Furthermore, the importance of clear and comprehensive job descriptions was stressed, to avoid misunderstandings. Alongside technical qualifications, participants emphasized the significance of soft skills such as teamwork, social skills, problem-solving, communication, and adaptability. While assessing a candidate's fit into the company culture and motivation level was deemed important, over-qualification was not seen as a sole exclusion criterion. Extraneous information, such as hobbies, relationship status, and photos, was considered irrelevant to job performance and was suggested to be excluded from initial screening.

Participants also discussed the importance of language and communication skills. As for the interview process, they suggested it should screen not only the technical competencies of the candidate, but also their soft skills, motivation, and cultural fit. Practical skills like forklift operation licenses and place of residence were also discussed depending on the nature of the job.

In contrast, some participants were against the requirement of time flexibility, suggesting such scheduling matters should be handled by managers. There were also differing opinions on the relevance of educational background, as some participants believe real-time work experience outweighs education.

When it comes to sensitive information such as pictures, opinions varied, although most agreed that priority should be given to education and professional experience. Other factors like distance from home to work, availability on weekends, knowledge of specific languages, and good health were also brought into consideration depending on the job role and its requirements. There was an agreement about issuing importance to these elements without creating a specific order. Overall, participants highlighted the need for a comprehensive, unbiased, and skill-focused approach towards candidate screening for a fair recruitment process.

Elimination Criteria

Below a broad list of the most relevant criteria identified by the participants during their profile analyses.

For the Assistant Store Manager position:

- ➔ Not meeting basic requirements
- ➔ Lacking educational background
- ➔ Overqualification
- ➔ Less strong past experience
- ➔ Legal regulations compliance in the specific Country
- ➔ Not eliminating women because of the DEI policy

Other elimination criteria that were discussed included the quantity and duration of prior experiences, whether candidates had not worked in the same sector as the job offer, disliked hobbies, and personal attitudes.

For the Logistic Officer position:

- ➔ Not meeting basic requirements
- ➔ Overqualification
- ➔ Less strong past experience
- ➔ Not being available for night-shifts
- ➔ Distance from work
- ➔ Not being flexible enough





Additional elimination criteria that were discussed included candidates having sufficiently short professional experiences, a weak educational background, and hobbies that could lead to inferences.

The elimination criteria proposed by different groups during the workshops varied significantly. While some found no reason to exclude anyone at that stage, others did identify certain candidates as less suitable.

For some, the absence of a career history in the FMCG (Fast-moving consumer goods) sector could serve as a disqualifying factor that influenced the perceived suitability of the candidate 3 (see [Annex 8](#)). Conversely, language skills and global experience weighed in favor of the first two candidates. Regarding the logistic officer position, someone raised concerns about logistics such as a candidate's residence location and any potential physical limitations.

A recurring theme was that of being overqualified. Perhaps surprisingly, some candidates were singled out multiple times for this reason, with the suggestion that they may not intend to remain in the assistant position they applied for. Similarly, other candidates were singled out for not having relevant experiences.

Some groups noted that biases towards gender and candidature inconsistent with the company's culture may serve as elimination criteria. A group recognized a bias towards female candidates due to the nature of working hours and travel, whereas some candidates faced possible elimination because of their nationality and potential relocation issues.

The influence of the given job description was discussed by several groups, some of whom expressed frustration regarding its vagueness and the problems this posed when attempting to screen out candidates.

Other factors such as educational background, social media activity, or inconsistency in job experiences were cited as potential reasons for exclusion. The diverse viewpoints throughout the exercise revealed different interpretations of the same information, manifesting an array of biases.

The interesting point with respect to elimination criteria was that all the points identified were not universally accepted and differed significantly among groups, highlighting the subjectivity inherent within the recruitment process.

Selection Criteria

Selection criteria were discussed starting from the relevant information extracted from the first question. Participant responses leaned towards fair play, comprehensive evaluation, and diversity. It was evident from the discussions that the majority of the groups found it hard to reach a unanimous decision on which candidate to interview based on the provided information.

A number of participants expressed their preference for a female candidate (candidate 2 – see [annex 8](#)), citing her experiences and education. Some suggested interviewing the candidate on account of the company's policy of supporting female employees. Others felt hesitant about this candidate, believing her extensive experience might over-qualify her for the position.

There was strong advocacy for a fair evaluation of all potential candidates beyond just qualifications and technical skills. This approach was underlined by many teams who suggested interviewing all candidates to gain an understanding of their soft skills, communication abilities and actual potential. The teams agreed on the importance of detailed, well-defined job offers in guiding the recruitment process.

Similarly, discussions arose around the aspect of a candidates' personal information. While some saw details such as parental status, age and nationality as potential sources of bias, others viewed them as crucial considerations for the job. For instance, a candidates' ability to relocate or their language proficiency were essential for some participants.





Concerns about the hiring of transgender individuals were also raised, emphasizing the need for alignment of the company's policies with the candidate's identity. This discourse further highlighted the relevance of diversity policies.

Judgments about candidates' motivation were scrutinized since these could not be derived from the given information. Simultaneously, certain groups valued the determination of candidates seeking higher education while in employment.

Key Findings

Here's a summary of the criteria considered for both the Assistant Store Manager and Logistic Officer positions. It's important to note that each group in every workshop operated independently. As a result, certain criteria were identified by multiple groups or in multiple workshops, while some points were only emphasized by a single group. Additionally, elements deemed crucial by some groups were considered preferable by others. The criteria listed below are those shared by multiple groups, with individual elements excluded.

Assistant Store Manager Position

Relevant Criteria:

- Meeting basic job offer requirements
- Relevant professional experience
- Duration of previous experiences
- Motivation and cultural fit
- Communication skills
- Education compensating for experience
- Knowledge of languages
- Availability on weekends
- Teamwork and social skills
- Legal regulations and DEI policy compliance
- Job description clarity
- Having a blind CV for initial screening
- Preferable factors: working in a similar sector, other soft skills, not being overqualified, time flexibility, hobbies, licenses, and training

Elimination Criteria:

- Not meeting basic requirements
- Lacking educational background
- Overqualification
- Less strong past experience
- Legal regulations compliance in the specific country
- Not eliminating women because of the DEI policy
- Other less important factors: quantity and duration of prior experiences, not working in the same sector, disliked hobbies, and personal attitudes

Logistic Officer Position

Relevant Criteria:

- Meeting basic job offer requirements
- Living close to the workplace
- Professional experience
- Educational background





- Night shift availability
- Flexibility
- Motivation and cultural fit
- Job description clarity
- Language and Communication skills
- Previous experience within the company
- Soft skills (teamwork, social skills, problem-solving, adaptability)
- Having a blind CV for initial screening
- Practical skills and specific job-related licenses
- Good health
- Other discussed factors: the relevance of time flexibility and educational background

Elimination Criteria:

- Not meeting basic requirements
- Overqualification
- Less strong past experience or lack of relevant experiences
- Not being available for night shifts
- Distance from work
- Not being flexible enough
- Additional criteria discussed: insufficiently short professional experiences, weak educational background, hobbies leading to inferences
- Elements of bias highlighted: potential gender discrimination, subjectivity within the recruitment process, cultural fit, interpretation differences in candidates' profiles.

Selection Criteria considerations (for both positions):

- Difficulty in reaching unanimous decisions on candidates to interview
- Preferences for fair play, comprehensive evaluation, and diversity
- Advocacy for the principle of job-relatedness (choice based on experiences and education)
- Hesitation about overqualification concerns
- Strong advocacy for fair evaluation beyond qualifications and technical skills
- Importance of interviewing all candidates to assess soft skills, communication abilities, and potential
- Emphasis on detailed, well-defined job offers guiding the recruitment process
- Debate on personal information's relevance, with some seeing it as a potential bias source
- Concerns about hiring transgender individuals and the importance of alignment with diversity policies
- Scrutiny of candidates' motivation and value placed on determination for higher education while employed.

In conclusion, the groups showed a preference for a candidate with well-aligned skills, experience, and qualifications, but they also underscored the importance of comprehensive interviews, well-defined job offers, and diversity policies in the recruitment process.

Relevance of the Company Rules

Several reports (5/7) from the workshops revealed that the company's policies were significant when deciding which candidates to invite or discard for an interview. Participants, and especially those with a legal expertise, felt that such policies provided a foundational framework that facilitated fair and





consistent candidate evaluation, thus playing a crucial role in eradicating bias or avoiding decisions based on whim.

The company's policies on supporting women were influential in choosing the most adequate candidates.

Some groups revealed that the company's policies shaped the HR managers' opinions, with candidate 1's company experience factoring into the selection process as he was previously recognized as an employee (see [annex 8](#)). There was an inclination towards hiring this candidate, given the company's policy to retain its current staff.

Again, some groups highlighted a lack of sufficient information about the company as a hindrance to their decision-making process. They were uncertain about the company's current diversity and language needs and whether these should impact their choice of candidates.

Participants highlighted the importance of potential employees adhering to company policies beyond mere rhetoric. Also mentioned was the significance of the company's diversity policy in promoting a trans person's candidacy.

A group highlighted the retention policy as a key determinant in candidate selection but revealed that opinions differed on how this translated into evaluation of candidates. In general, this was a discussed element in different groups: some suggested a family-oriented candidate, while others benchmarked an ambitious candidate as the ideal choice. In some instances, however, policies on prioritizing those with prior company experience or international exposure did not carry much weight. Overall, the company's rules and policies emerged as crucial components in the recruitment process. The strategic importance of providing feedback to those not invited for an interview was also indicated, as it promotes transparency and builds trust in the hiring process, thus enhancing the employer's reputation.

Consensus Issues

The discussion among the HR officers exposed diverse and often conflicting views on various aspects of the hiring process. While in some areas a shared perspective emerged, the overall consensus was not always unanimous. The reflection ranged from candidate selection criteria to the interpretation of personal, academic, and professional information, to matters like objectivity, company policies, and more.

One intriguing subject was the introduction of industry-specific experience in different sectors. Some participants viewed it as a likely problem, while others did not. The issue of candidates being assigned based on their place of residence initially met with mixed rates. However, participants later reached a common understanding that any applicant would be considered. Personal attributes deduced from hobbies also stirred discussion, but no consensus was reached.

Another assessment revolved around the inclusion or exclusion of personal information - mainly children's status in CVs. The biased perception between men and women having kids and its reflection on their career and job eligibility sparked an emotional outpour in some groups, but concluded with no definite resolution.

Divergences were also noticeable on the importance of a detailed job profile, adherence to company rules, feedback for unselected candidates, prioritization of qualifications versus motivation, and the provision of all GDPR information. Participants also dwelled on the significance of interviews, particularly when company policies complicated the selection process.

The biggest discrepancies were in specifying marital status, the impact of nationality and language competence, and quantifying personal information in recruitment technology. Some were critical of considering irrelevant information like marital status and children, calling it a cause for discrimination. Strikingly, no comments were made regarding one candidate being non-binary.





While a general understanding was reached in some areas, complete consensus was not always achieved. However, the broad consensus pointed towards the importance of job interviews for better selection: regarding criteria and factors that remained ambiguous or where consensus was elusive, the preferred strategy suggested is to proceed with candidate interviews to clarify the issue or delve deeper.

Plenary Discussion

The activity showed a broad range of perspectives.

HR experts pondered on the fine line between affirmative action and bias. They concluded that all candidates deserve a fair interview and to be judged based on merits, not their demographic backgrounds. They also noted difficulties in decision-making due to a lack of company information.

Legal experts emphasized the importance of adhering to company policies to avoid potential legal problems. They pointed out that failure to invite a qualified candidate to an interview might pose a legal issue if the company's procedures had not been correctly followed. Legal experts also highlighted the inherent privacy risks and potential for discrimination linked to personal and sensitive data processing. They mentioned the "Privacy by design" concept, which emphasizes the importance of prevention over correction, the ensuring of security throughout the product or service lifecycle, and user-centrality. Legal experts also pointed out the potential role of an AI system in accentuating discrimination.

In contrast, AI experts drew attention to the high turnover rate among women in IT, suggesting the need to recruit more women simultaneously to address the issue. They also suggested considering job relatedness and gender quotas directly during the development of the recruitment software and introducing a matrix to ensure recruitment rules adaptability. Instead of advocating for a system that outright excludes candidates, they suggested implementing a ranking system to highlight profiles that merit additional consideration or pose potential issues. The conversation then delved into questioning the feasibility of preventing bias based on data: despite efforts to create a fair system in terms of algorithms, evaluation metrics, and processes, the challenge persists due to candidates providing data that may introduce bias.

NGO representatives stressed the significance of considering various factors like race and sexuality while hiring. They highlighted the risk of minority exclusion and the problem of making assumptions about language competence based on nationality.

Finally, individuals with a philosophical background offered a critical reflection on the risk of negating subjectivity in the recruitment process, emphasizing the difficulty in evaluating soft skills objectively. They suggested skill evaluations without bias-oriented discrimination.

Philosophers also raised issues about unconscious bias and the subjective nature of objectivity. They proposed robust guidelines for consistency and questioned how HR professionals define objectivity, suggesting that quantifying it might also lead to bias.

10.2.3 Discussions into groups: takeaways

In general, the discussions proceeded smoothly, revealing numerous interesting and pertinent insights. Although a broad consensus was often not achieved, participants expressed satisfaction with the topics. It is important to note that reaching a general agreement was not the primary objective of these workshops. Instead, the focus was on extracting critical information and opinions on AI concepts in recruitment, particularly concerning the two presented tools. The introduction of the Debiasser provided valuable insights into participants' initial understanding and perspectives. This was particularly interesting as it allowed for an initial analysis of how to effectively explain the tools and gather feedback from experts regarding their integration.

The first and second activity discussions provided rich insight into the myriad aspects of fairness in the recruitment process. Participants agreed that fairness should equate to providing a level playing field for





all applicants, with emphasis placed on objectivity, consistency, and the elimination of bias. Professionals' involvement and the use of AI tools were suggested to achieve a more equitable recruitment process, thus revealing the ambiguity around how AI tools are perceived, as both potential sources of bias and instruments to counter them. One key point emphasized by many participants in the workshops was that the concept of fairness, as presented in the activity, becomes fragmented when considering only one element. Objectivity, consistency, non-manipulation, job-relatedness, non-discrimination, and multiple perspectives are key aspects that cannot be separated from each other to ensure a fair process.

The third activity prompted participants to analyze and structure the logical reasoning behind a hiring decision, specifically identifying critical criteria such as selection factors, elimination factors, inclusion of company policies, and other relevant considerations in the delicate process. It's important to note that not every participant engaged with the same level of involvement; the HR experts played a key role in initiating the discussion, while legal, philosophical, and AI perspectives provided diverse insights into solutions for these dilemmas. Altogether, the discussions highlighted the need for a dynamic balance between diverse considerations in the recruitment process. The feedback indicated a need for a multi-pronged approach, ensuring fairness, inclusivity, privacy, and transparency in recruitment procedures.

10.3 Interactive/hands on work: which requirements for AI tools in recruiting?

10.3.1 Feedback from the rapporteurs

The interactive segment of the workshops involved identifying requirements for the AI tools. In the subsequent reports, the rapporteurs were tasked with evaluating how the activity unfolded. This assessment included gauging what aspects were perceived as positive or negative, as well as determining participants' overall satisfaction with and utility of the activities. The group work proceeded in a generally positive way, allowing for a multitude of insights and valuable discussions. However, certain concepts, especially the complexities of the AI tools in debiasing recruitment and the poster organization, proved difficult for some participants, particularly non-experts, to comprehend. The first activity entailed understanding the principles of NLP and CBR, which proved challenging to non-experts, even if they were briefly outlined in the first part of the workshop. Despite this, there was a noticeable eagerness to learn. The second activity, focused on identifying requirements for the AI tools, was easier for participants, with HR managers and AI experts leading the discussions. It shall be noted that participants based their observations and suggestions on requirements for AI tools on very different imageries and understandings, often even misunderstanding the specific use of AI applied to selection and recruitment, not necessarily in line with the Debiaser technological development plans/methodology.

The distribution of participants in the groups created rich discussions, even though not all participants always agreed. In particular, the groups with a higher number of HR representatives saw more complex discussions due to the emergence of diverse perspectives. There was a smooth collaboration among the different stakeholders, with non-AI-experts such as HR managers and legal experts actively contributing their ideas and taking the opportunity to understand the technical feasibility of their requirements.

It's important to note that the entire workshop was specifically designed as a brainstorming activity, intending not to reach a consensus but to stimulate discussions, identify potential issues, and explore collective solutions. The primary objective was to initiate conversations that would later be subjects of more in-depth analysis during the international workshop in December 2023.

10.3.2 The perceived needs and requirements for AI tools

The participants classified the identified requirements for AI tools into the three phases of the recruiting process. The initial brainstorming sessions focused on needs that an AI technology such as the Debiaser could/should meet, how it could support the selection and recruitment process in its 3 suggested phases:





- Phase 1 - Ticking off the basic/must-have requirements
- Phase 2 - Scanning for preferred or good-to-have qualifications
- Phase 3 - Matching the holistic picture of the candidate to the role

Brainstorming

During the brainstorming sessions, the participants' focus was predominantly on the general use of a tool that would encompass the entire recruiting process. Many ideas were shared with the concept of an intelligent robot performing the duties of a recruiter, although this was never presented as the intention of the consortium. Afterwards, it was asked to deep dive into each phase and identify the specific needs of an AI tool in terms of functional requirements, input and output expected and conditions to ensure fairness.

The **emerging needs** can be clustered under seven distinct topics, presented in the below table in a schematic format, highlighting those that are more relevant for the Debiaser NLP (in purple), for the CBR models (in green) or for both (in yellow).

Table 29 Emerging needs for the NLP based and CBR Debiaser tools

	CBR	NLP	
Business Needs	Fill vacant positions in the organization	-	
	Optimal matching between candidates and organization		
	Simplifying the overall hiring process		
	Reduce the time of application screening		
	Make the comparison of diverse job applications easier		
Screening Phase	Different outputs for each recruitment phase	AI cannot make the choice but can serve as a supporting tool	
	Integrate and synchronize the outcomes of phases 1 and 2		
	Additional phase 4 to record information on final decisions and monitor potential bias patterns across all hires		
	Extra phase (Transversal Phase) for compliance with policies and regulations		
	Restricting AI involvement to only the first phase		
System Input	Use of pre-structured CV provided by the company	Include the social profiles as input data	
	No information extraction from a text written by candidate: standardized blind CV formats		
	Cover letters and resumes as the primary inputs		
	"Killer questions" as supplementary inputs alongside the CV		
System Output	Generation of a shortlist for interviews as the primary output	-	





	CBR	NLP
	No elimination, only preference stating	
	Identify the appropriate compensation benchmark based on experience, market trends and company needs	
	Identify alternative requirements that may lead to a different candidate profile with similar skills	
System Usage	Ensure training programs for the design and implementation of the technology	
	Intuitive and easy to navigate interface	
	Ensure efficiency, effective usage, comprehensible structure	
	Adaptability to the specific regulations of the different countries (regional nuances)	
Model Training	-	Ensure a diverse range of demographic samples in the training data
		Maintain training data that is devoid of discriminatory patterns and ensure regular updates
		Have a significant volume of training data
	Adaptive Learning: feedback to continuously learn	
	Cross validation across different use cases	
Ensuring Fairness	Identify essential skills and exclude superficial attributes	-
	Compliance with data protection regulations	
	Refrain from emphasizing or considering irrelevant elements that might introduce bias	
	Precision and ability for supporting 'innovative' thinking	
	Implement periodic reviews, inspecting data sources for signs of bias	
	Explainability and transparency through effective programming	
	User awareness: inform the candidate about the involvement of AI in the recruitment process	

Here are the results of the brainstorming activity explained in more details:

- **Business Needs:** identification of requirements aligning with the overall business objectives and goals related to the recruitment process.
 - Addressing the need to efficiently fill vacant positions within the organization [CBR]
 - Ensuring optimal matching between candidates and available positions [CBR]





- Simplifying and streamlining the overall hiring process for improved efficiency [CBR and NLP]
- Enhancing HR efficiency through improved and timely screening of incoming job applications [CBR and NLP]
- Facilitating the comparison of diverse job applications for more effective decision-making [CBR and NLP]
- Screening Phases: considerations and suggestions around each phase of the recruiting process, screening, evaluation, and selection.
 - Tailoring outputs based on the unique requirements and objectives associated with each phase of the recruitment process [CBR]
 - Integrating and synchronizing the outcomes of Phases 1 (Ticking off the basic/must-have requirements) and 2 (Scanning for preferred or good-to-have qualifications), with candidates successfully passing both phases receiving higher scores [CBR]
 - Including an additional Phase 4 to record information on final hiring decisions and monitor potential bias patterns across all hires [CBR]
 - Introducing an extra phase (Transversal Phase) to monitor compliance with various policies, regulations, and privacy considerations [CBR and NLP]
 - Restricting AI involvement to the initial phase, with only the first phase choice being determined by the AI [CBR and NLP]
 - Specifying that AI cannot make the Phase 3 choice (Matching the holistic picture of the candidate to the role) but can serve as a supporting tool in the decision-making process [NLP]
- System Input: exploration of requirements related to the data and information provided to the system as input.
 - The use of a pre-structured CV provided by the company as a primary source of information [CBR]
 - Not extracting information from a text written in a format chosen by the candidate, emphasizing the need for standardized CV formats [CBR]
 - Introducing additional "killer questions" specific to a role, which would serve as supplementary inputs alongside the CV [CBR and NLP]
 - Designating cover letters and resumes as the primary inputs for the AI [CBR and NLP]
 - Including the social profiles of candidates as additional input data [NLP]
- System Output: identification of requirements concerning the presentation and delivery of the system's outcomes.
 - Emphasizing the generation of a shortlist of candidates deemed suitable for interviews as one of the primary outputs [CBR]
 - Stating a preference for the AI system to focus on identifying preferred candidates without outright eliminating others [CBR]
 - Specifying that the AI should identify the appropriate compensation benchmark for each candidate based on their experience, considering market trends and company requirements [CBR]
 - Highlighting the need for the AI to identify alternative requirements that may lead to a different candidate profile while maintaining similar skills, deviating from the original job description [CBR and NLP]
- System Usage: consideration of requirements related to the practical utilization of the system by recruiters.
 - Ensuring training programs to accompany the design and implementation of the technology, ensuring that users are well-versed in its functionality [CBR and NLP]
 - Prioritizing the development of an interface that is intuitive and easy to navigate for users interacting with the AI system [CBR and NLP]





- Stressing the importance of the AI system being straightforward and easily comprehensible for users, promoting efficient and effective usage [CBR and NLP]
- Adaptability of the AI system to the specific requirements and regulations of different countries, acknowledging regional nuances [CBR and NLP]
- **Model Training:** exploration of requirements associated with the training of the underlying models.
 - Adaptive Learning: receiving feedback from recruiters to continuously learn and improve its performance over time [CBR and NLP]
 - Ensuring that the outputs of the AI model are rigorously validated across different use cases to verify its effectiveness and fairness [CBR and NLP]
 - Ensuring that the training data for Natural Language Processing incorporates a diverse range of demographic samples to prevent the reinforcement of societal biases [NLP]
 - Maintaining training data that is devoid of discriminatory patterns and ensuring regular updates to promote fairness in the AI system's outputs [NLP]
 - Significant volume of training data is needed to prevent the replication of errors and biases in the AI system's decision-making [NLP]
- **Ensuring Fairness:** focus on requirements addressing the need for fairness in the recruitment process.
 - The AI should selectively identify and present the essential skills and main technical characteristics of the candidate, excluding superficial attributes [CBR]
 - Compliance with data protection regulations to ensure the lawful and ethical use of candidate information [CBR and NLP]
 - The tool should refrain from emphasizing or considering irrelevant elements that might introduce bias into the decision-making process [CBR and NLP]
 - Enhance the tool's precision and ability for supporting 'innovative' thinking across all three phases of the recruitment process [CBR and NLP]
 - Implement periodic reviews of the tool, inspecting data sources for signs of bias or unfair weighting to ensure ongoing fairness [CBR and NLP]
 - Improve the tool's explainability and transparency through effective programming, allowing users, particularly recruiters, to understand the decision-making process [CBR and NLP]
 - Transparency with candidates by providing information about the involvement of AI in the recruitment process [CBR and NLP]

Despite the divergence from the intended focus on the CBR and the NLP systems, these seven clusters provided a valuable framework for understanding the multifaceted requirements perceived by the participants.

Breaking down the needs: phase 1,2,3

In the second part of the exercise, participants were asked to identify requirements within the three phases of the recruiting process. Reflecting on the outcomes of the exercise and the feedback provided by the groups at the conclusion of the workshop cycle, it became evident that the two tools, the Case-Based Reasoning (CBR) system and the Natural Language Processing (NLP) system, were perceived with distinct strengths. The CBR tool was deemed more suitable for the first phase of screening, while the NLP-based system emerged as a versatile, cross-phase tool, with particular efficacy during the third phase. This is the phase where recruiters aim to achieve a holistic match between a candidate and the company, as it is at this stage that recruiters are more susceptible to potential biases, having to carefully analyze all the textual information about candidates.

In the three phases, the key topics identified are:

- Business needs of the specific phase (as above)
- System input (as above)
- System output (as above)
- System usage (as above)





- Ensuring Fairness (as above)
- Useful Tools: proposals of solutions for specific requirements

Phase One - Ticking off the basic/must-have requirements (mostly related to CBR)

Business Needs:

- Streamline initial screening to reduce errors due to rushed decision-making.
- Identify candidates with non-negotiable priorities.

System Input:

- Extract information from a pre-structured CV provided by the candidate during the application phase.
- Avoid extracting information from texts in formats chosen by candidates to minimize bias risks.
- Incorporate additional details from standardized questionnaires.
- Use a blind CV format, including essential details such as education, visa status, languages, licenses/certificates, hard skills, management skills, and experiences.
- Utilize detailed job descriptions.

System Output:

- Verify if candidates meet the essential requirements of the job.
- Rank candidates based on objective criteria.
- Identify desirable qualities outlined in the job description.
- Generate a compliance report comparing candidates against the job description.

System Usage:

- Implement a user-friendly database for easy access to candidate information.
- Include filters to visualize candidates based on common variables.

Ensure Fairness:

- Exclude superficial or biased characteristics to prevent subjectivity.
- Enable the tool to interpret various CV formats.
- Ensure consistent output for similar profiles.
- Obtain comprehensive consent from candidates for handling personal information.

Useful Tools:

- Visualization grid.
- Evaluation matrix.
- Deterministic model using EITHER/OR logic for matching keywords in CVs.
- Supervised qualification with labels IN/OUT.
- Keyword extraction tool for skills identification.
- NLP text cleaner for refining and processing textual data.

Phase Two - Scanning for preferred or good-to-have qualifications (mostly related to CBR)

Business Needs:

- Identify candidates with multipotential.
- Differentiate between must-have and nice-to-have requirements.
- Verify key requirements in the job offer.
- Focus and prioritize information for the decision-maker [NLP].

System Input:





- Output from Phase 1 (matching skills and must-have requirements).
- Ranking of preferable candidates.
- Evaluation of IT skills, organizational skills, communication skills, accuracy, and languages (based on job relevance).
- Candidate's salary expectations.

System Output:

- Ranking of preferable candidates.
- Explanation of the choices made by the system.
- Ability to filter candidates based on specific requirements.
- List of "neutralized" candidates, cleaned from biases [NLP].
- Suggest questions for HR representatives to ask in potential job interviews [NLP].
- Identify missing information essential for evaluation [NLP].

System Usage:

- Implement a user-friendly database for easy access to candidate information.
- Include filters for visualizing candidates based on common variables.

Ensure Fairness:

- Ensure an adaptive and continuous learning process.
- Ensure consistent output for similar profiles.
- Support recruiters by identifying complementary requirements in case the job description is biased.

Useful Tools:

- Evaluation matrix.
- Unsupervised ranking.
- Framework for analyzing and comparing CVs.
- Implement explainable AI features.

Phase Three - Matching the holistic picture of the candidate to the role (mostly related to NLP)

Business Needs:

- Identify mismatches between policies and workplace culture.
- Focus and prioritize information for decision-makers.
- Identify multipotential candidates.
- Ensure compliance with privacy policies.

System Input:

- Hiring decisions for adapting learning.
- Output of the previous phase 2.
- Primarily soft skills assessment.
- Complete CV and cover letters of candidates (without restrictions).
- Company information, values, descriptive documents, policies, diversity plans.

System Output:

- Shortlist of candidates for interviews.
- Assessment of heterogeneity/homogeneity of the selected shortlist.
- Fitting or matching score (%) for each candidate based on multiple metrics.
- Identification of potential bias as red flags.
- Highlighting certain CV information with further explanation of interest.





- Support compensation formulation based on market needs.
- Manipulation of CVs to ensure bias-free applications.
- Identify missing information essential for evaluation (objective requirements).
- Extract relevant information from cover letters.
- Add university and department information with relevant topics or categories.
- Highlight keywords in cover letters relevant to the job offer.

System Usage:

- Final decision made by HR specialists.
- System should be trustworthy for HR specialists.
- Implement a Human Feedback Loop with UI for human recruiters to provide feedback.

Ensure Fairness:

- Multi-level evaluation based on different variables (multiperspective).
- Set metrics for evaluation to objective criteria.
- Provide explanations for every decision.
- Implement adaptive learning to receive feedback from recruiters.
- Put countermeasures in place to mitigate bias.
- Avoid negative reactions to factors like "overqualification".

Useful Tools:

- Visualization grid.
- Customizable and context-sensitive threshold of acceptable candidates.
- Use multiple scores and minority belonging checkboxes for the final shortlist.
- Implement a chatbot for user interaction.

The concluding exercise of this workshop cycle revealed a notable divergence in ideas and perspectives among participants. To provide more clarity, the following is a summary of the divergent positions and ideas that emerged during this exercise.

Differences arose on whether AI can engage in positive discrimination and the effectiveness of this approach. On a related note, some participants disagreed on the use of legally mandated hiring quotas for individuals with different disabilities. While some saw it as a form of positive discrimination, others viewed it as bias and promoted the idea of competency-based hiring.

There were divergent views on whether AI could fully manage the recruitment process, with some participants suggesting that it should screen applications and help with decision-making, while others proposed a more holistic role for AI, encompassing the entire recruitment process.

Participants disagreed on how to measure diversity and whether it should be reflected by the AI tool. There was a discussion on whether an AI tool should move away from typical CV applications towards a form-based approach to capture job-related information better. Some argued the AI tool should merely aggregate and filter data to support decision-making, but not make the decision itself.

Participants had divergent opinions on the role of AI in hiring processes. Philosophers highlighted the risk of bias at the holistic match stage, while lawyers emphasized the importance of following privacy laws. HR representatives expressed skepticism of AI's efficacy and reliability in the recruitment process, requiring human oversight.





Participants expressed differing views on how far AI should go in the recruitment process, with discussions on whether the output should mostly be based on the CV in the first phase. AI developers defended the technology while HR representatives were more cautious.

Finally, divergences arose regarding whether personal characteristics that could lead to biased outcomes should be eliminated by the AI tool. Participants also disagreed on whether to include more personal experiences among the inputs, with some citing potential discrimination and others arguing that personal experiences enrich the candidate profile.

10.3.3 Interactive work - Challenges and Takeaways

Overall, the activities' complexity did slow things down somewhat. The poster's structure proved to be difficult to decipher for some and the use of sticky notes, while intended for organizing the emerging content, in certain cases ironically slowed the conversation. The most successful activity of this interactive session was the final evaluation (walking plenary), where all participants easily expressed their thoughts and perspectives. Some groups faced challenges in completing the exercise, occasionally leaving the third or fourth activity unfinished. The individual partners' reports underscored the importance of enhancing clarity in poster designs and streamlining note-taking processes for future workshops

When delving into the functional and business requirements for an AI tool, participants appeared to struggle with understanding the actual functionalities and use of the two systems. Despite the prior presentation and explanation of the two tools before the practical activity, groups encountered difficulties in comprehending the intricacies of the explained tools. It's worth noting that the level of AI knowledge required to understand such tools is advanced, and not everyone was expected to have a deep understanding of how they function. Nevertheless, particularly during the brainstorming activity, participants did not focus on how the specific NLP and CBR systems would be integrated into the phases of the recruiting process. Instead, their attention shifted towards discussing the ideal AI tool they would like to have to support the recruiting process, leading the activity and discussions out of scope.

The focal point emphasized during this workshop highlighted the importance of participants gaining a deeper comprehension of the Debiaser's functionality and application. It was stressed that participants should observe practical demonstrations and actively engage in hands-on exercises to enhance their understanding and proficiency with these types of tools. This aspect was the main focus when developing the methodology for the first international co-creational workshop, in Venice.

Key Takeaways:

- Although the exercise was perceived as complex in its nature, people participated with enthusiasm and positivity, bringing up relevant questions and new discussions
- Often participants had different perspectives and could not reach a consensus over some thematic: particularly, there was often a divergence on the role of positive discrimination and hiring quotas for individuals with disabilities.
- Opinions on the extent of AI involvement in the recruitment process varied, with key decisions emphasizing AI as a supporting tool rather than a stand-alone solution. Decisions included limiting AI access to sensitive data, excluding it from the decision-making process, and transforming it into a tool focused on identifying strictly positive factors.
- Participants frequently speculated that decisions based solely on CVs were risky, emphasizing the need for an AI system to incorporate additional elements to ensure fairness, objectivity, and non-discrimination. Additionally, there were differences in views on eliminating personal characteristics and including personal experiences in AI-driven recruitment processes
- Disagreements frequently arose regarding the measurement of diversity, with discussions often dominated by how to handle and ensure diversity. Establishing a common understanding proved challenging for participants.





- The risk of bias was found hard to measure, because it was pointed out that it is hard to measure even in human minds. Both philosophers, lawyers, and HR representatives pointed out the fundamental importance of regulation, privacy law and AI efficacy in environments that must guarantee fairness.

In conclusion, while the workshop saw several divergent views on various aspects of bias in the AI recruitment process, there was a consensus that AI should assist, not replace, human decision-making in the process.





11. The International Co-Creation Workshop- Methodology

11.1 Workshop’s agenda and target

The International co-creation workshop was held on December the 7th, 2023 in Venice. This workshop had different purposes:

- Collecting constructive feedback through engagement with a simulated tool designed to replicate the future CBR system and word-embedding based system envisioned for the project.
- Encouraging participants to contemplate the concept of trustworthiness in AI systems and delineate system requirements in alignment with the ALTAI paradigm of trustworthiness in AI.
- Addressing the need to develop effective training packages by gathering ideas and input on learning requirements.

The workshop saw 45 people registered and 40 who participated. Among these people there were stakeholders, representatives, and partners, who would work as facilitators, rapporteurs or passive participants during the workshop’s activities. The composition of the guest participants was as follows:

Table 30 External stakeholders at International Co-Creation Workshop

Country	Role
EE	Worker/Worker representative
EE	Advocacy organization/NGO
EE	HRM employee
IS	HRM employee
IS	AI Developer
IS	AI Developer
CH	HRM employee
CH	Worker/Worker representative
IT	HRM employee
IT	HRM employee
IT	AI Developer
IT	Worker/Worker representative
NL	HRM employee
NL	Advocacy organization/NGO
TR	Advocacy organization/NGO
TR	AI Developer
NO	HRM employee
NO	Advocacy organization/NGO
NO	HRM employee





The following number of project members from each consortium member were present:

Table 31 Project members from each consortium member

Partner	Workshop Participants
NTNU	7
HI	3
SVEN	4
ULEID	2
FARPL	2
BFH	2

The event lasted approximately 1 day (8 hours) and was structured as follows:

Table 32 Programme of the International Co-Creation Workshop

Time	Programme
9:00 - 9:30	Arrival and registration
9:30 - 10:00	Welcoming, introduction and Warming up exercise to familiarize with each other
10:00 - 11:00	Simulating the use of AI based systems in recruitment and Selection_ 1st part Prioritizing features and Ranking candidates
11:00 - 11:15	Coffee break
11:15 - 12:00	Simulating the use of AI based systems in recruitment and Selection_ 2nd part Identifying and mitigating BIAS from CVs and cover letters
12:00 - 12:45	Plenary discussion
12:45 - 14.15	Lunch break
14:15 - 15:00	Trustworthiness & fairness: a brainstorming session
15:00 - 16:00	Exploring learning needs and expectations from capacity building
16:00 - 16:15	Wrap up, thank you and closing the meeting

11.2 First Block - Simulating the use of AI based systems in Recruitment and Selection

The initial half-hour of the workshop was dedicated to welcoming participants and introducing the BIAS project. To facilitate engagement and build a sense of camaraderie, an ice-breaking game was organized. Participants were grouped by country of work and invited to briefly introduce themselves. Each group, while holding a tossed-around teddy bear, shared one positive bias and one negative bias about their respective countries. This interactive activity aimed to create a relaxed atmosphere, allowing attendees to become acquainted with the surroundings, feel at ease, and gradually delve into the workshop's theme.

The first block of activities focuses on simulating human-machine interaction in the hiring process, namely the simulation of an AI Biased tool supporting selection with de-biasing functions. In this way, participants had the opportunity to become acquainted with and explore two distinct tools: the candidate ranker and the mitigation tool. These tools serve as simulations, with the candidate ranker emulating the CBR/NLP system for selecting suitable candidates, and the mitigation tool aims to simulate the word embeddings based system that identifies potential biases during profile screening. These exercises aim to gather people's feedback on the tools and establish specific requirements and details. Participants had previously been provided with a real-life scenario: they had received a job advertisement, a list of features, and some questions via email, and had the opportunity to reflect on them. Starting from their considerations, participants were then asked to configure the tools and evaluate their functionalities.





How do the simulation tools work?

Underneath the tool interfaces, Chat-GPT4 was employed: ChatGPT is a language model developed by OpenAI based on the GPT-3.5 architecture. GPT stands for "Generative Pre-trained Transformer," and it is a type of artificial intelligence model designed for natural language processing tasks. In the case of ChatGPT, it is specifically fine-tuned for generating human-like text responses in a conversational manner. The model is trained on a diverse range of internet text and has the ability to understand and generate coherent and contextually relevant text. When given a prompt, ChatGPT uses its pre-trained knowledge to predict what words should come next in the sequence. It doesn't have a deep understanding of the semantic meaning of the text in the way humans do, but it has learned statistical relationships and patterns from a vast amount of data. This allows it to generate coherent and contextually relevant responses that are often grammatically correct and contextually appropriate.

The model works by processing input text in a sequential manner, updating its internal state as it encounters each word or token. It maintains a contextual understanding of the input, allowing it to generate responses that are influenced by the preceding context. However, it's important to note that the model does not possess true comprehension or consciousness; it's essentially predicting what comes next based on its training data. Users can leverage this predictive capability for a variety of natural language processing tasks, such as conversation, question answering, summarization, creating conversational agents, generating human-like text and more. While the model excels at generating human-like text, it may not always produce perfectly accurate or contextually appropriate responses, and its limitations should be considered when using it in different applications. Participants were warned that the simulation tools are based on this technology and that ChatGPT could "hallucinate," meaning that it could potentially invent facts, realities, and information that may not be accurate or based on real-world data.

The tools share similar structures, differing mainly in user input, the case-specific instructions, and final output. Users interact with a browser-based web app, where they input or select data. In the backend, user input combines with a predefined prompt containing all relevant information for processing the output. This includes the job advertisement and the candidate profiles, previously provided via email and printed on each group's table for accessibility. The prompt also features specific instructions to guide the system toward the desired outcome. When the prompt is generated, the app sends the API (Application programming interface) call to ChatGPT and receives the output as a text file, which is displayed in the interface and made available to the user.

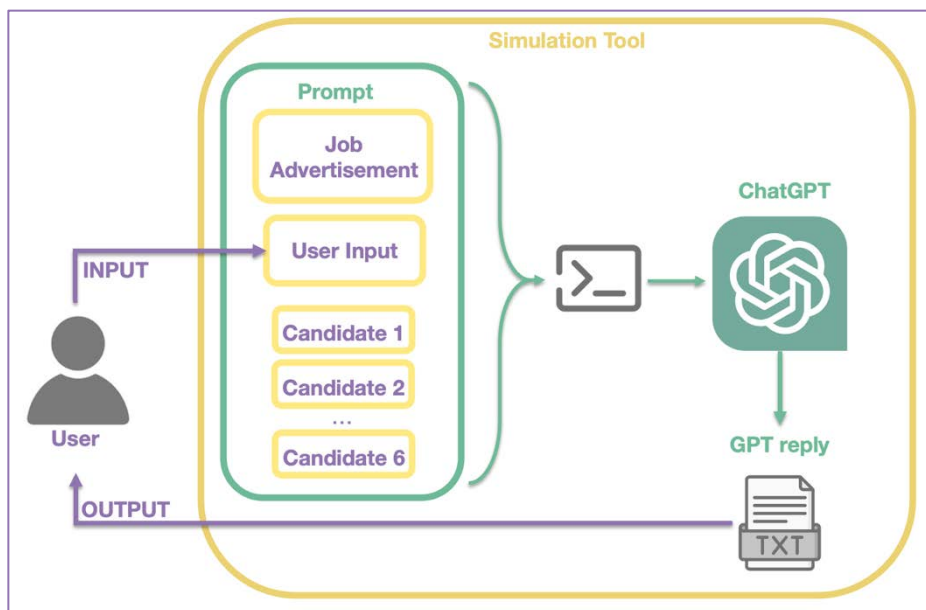


Figure 18 Interactions between a simulation tool and the user

Groups composition:

- Group 1 - **APRICOT**:
 - 1 AI Developer
 - 1 HR Specialist
 - 2 Advocacy organization/NGO
 - 2 Partners
- Group 2 **BANANA**:
 - 1 AI Developer
 - 2 HR Specialists
 - 1 Advocacy organization/NGO
 - 2 Partners
- Group 3 **CHERRY**:
 - 1 AI Developer
 - 1 HR Specialist
 - 1 Advocacy organization/NGO
 - 4 Partners
- Group 4 **LEMON**:
 - 1 AI Developer
 - 1 HR Specialist
 - 1 Advocacy organization/NGO
 - 3 Partners
- Group 5 **PEACH**:
 - 1 AI Developer
 - 2 HR Specialists
 - 1 Advocacy organization/NGO
 - 2 Partners
- Group 6 **MANGO**:
 - 1 AI Developer
 - 1 HR Specialist
 - 1 Advocacy organization/NGO
 - 4 Partners

11.2.1 The Candidate Ranker: prioritizing features and Ranking Candidates – Methodology

Overview:

Participants were asked to focus on the candidates' ranking phase within the recruiting process, where they were tasked with identifying the most pertinent evaluation features, desirable elements, as well as those deemed irrelevant.

Afterwards, participants were tasked with interacting with the Candidate Ranker tool, a service that works as an automatic ranker for different profiles, based on their CV, their cover letters, and the relevant features identified by the groups. Participants were required to share their feedback on the experience and underline criticalities and opportunities for such a system. The purpose of the exercise was to examine the distinctions between human decision-making and machine decision-making and determine whether there were any common elements. For every group there was a facilitator, who was a Consortium representative responsible for coordinating the discussion and addressing the questions, and a





rapporteur, who was likewise a Consortium representative but was rather responsible to take notes and report the collective answers in the given template. Even if there were different questions in multiple sections of the exercise, it was asked to facilitators not to fixate on answering every single question, but rather try to cover each section using the questions as an overall guide.

The “Candidate Ranker” functionalities:

The simulation tools incorporate ChatGPT as the reasoning component, aligning with the primary goal of encouraging participants to engage actively with AI tools. To facilitate seamless interaction and enhance task handling, a simple interface has been created using Anvil. Anvil, based in Cambridge, is a web development environment aimed at empowering developers to quickly create and deploy web applications. Anvil provides an integrated environment with features like database management, task scheduling, secure user authentication, and easy integration with various external services through Python SDKs; developers can design the user interface with drag-and-drop functionality and code entirely in Python.

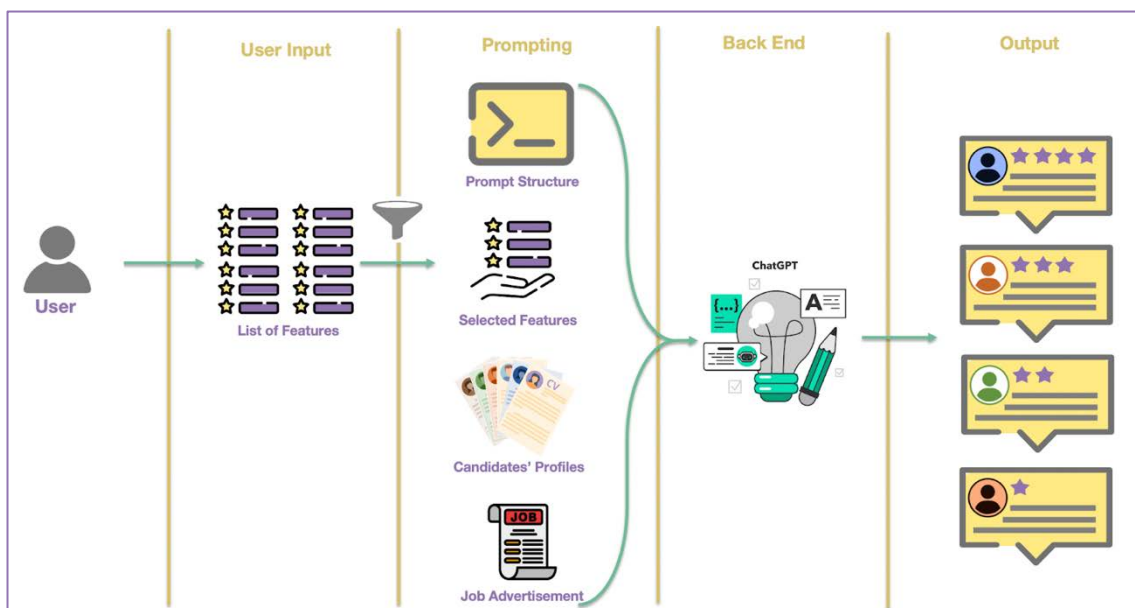


Figure 19 Simplified User Journey with the Candidate Ranker Tool

Both the simulation tools are designed in a way that all the needed data (e.g., candidate’s profiles, job advertisement, cover letters) are already integrated into the system storage, and there is no need to prompt any request by users, it all happens behind the scenes. In the first simulation tool, the “Candidate Ranker”, the user interacts with the system by choosing from a list of pre-selected features the ones that are most relevant in their opinion. The selected features, together with the other input data (candidates’ profiles including CV and cover letters, job ad, text instructions and request), would then compose the final prompt, which is sent to the ChatGPT API linked to an OpenAI account. In such a prompt it is asked to ChatGPT to act like a recruiter who has to rank the candidates based on the features selected (instructions). Finally, the output generated by ChatGPT is saved as a text file and shown to the user through the web application designed with Anvil. For clarity, here is an example of a simple prompt structure - including all the instructions - that is sent to the API (in the simulation a slightly more complex one was used). Also note that the prompt can be either simplified or made more articulated, depending on the requirements of the simulation.

Here is a job ad:
[Job_Adversiment]

Here are the following candidates, with cover letters and resumes:
[List_of_Candidates]



**Instructions:**

Assign a job suitability score on a scale of 100 to each of the candidates and rank them.

The job suitability score must be based on the following features evaluation:

[List_of_Selected_Features]

Present the candidates in order of ranking and provide an explanation for the ranking and the score. Every candidate must be presented as follows:

Candidate: ...

Ranking: ...

Explanation: ...

Total Suitability Score: ...

The exercise:

This exercise involves individual reading of candidate profiles, focusing on strengths, weaknesses, and potential biases (this was for the second exercise) in their CVs and cover letters. This task should be performed individually before the workshop by the participants, who received the material via email the day before. In this specific simulation, uniform data was provided to all participants, consisting of a job offer for the role of "Assistant Store Manager/Salesperson," along with detailed information about the company "XYZ." The participants received comprehensive details on six candidates, encompassing their resumes and cover letters. The candidates included Anna, Diego, Felix, Mark, Mohammed, and Priya. Notably, the creation of the simulation data was meticulously crafted to ensure a broad range of characteristics and biases were represented. This diversity covered aspects such as gender, disabilities, cultural background, and more. The profiles of all candidates are available at the [Annex 14](#), while the Candidate Ranker guidelines, and the detailed technical instructions to use the tool are available at the [Annex 15](#).

During the workshop, team members are asked to discuss and prioritize features for configuring the Candidate Ranker tool: the configuration consists in selecting an arbitrary number of features, and for each one of them, setting the priority level, which can be high (must-have feature) or low (nice-to-have feature); in the application there are up to five fields for the features, but users can add more through the text box at the bottom. The pre-listed features were extracted starting from the exercise of the second round of workshop ([activity 7.2.3](#)) where the participants were asked to elaborate on the selection and elimination criteria in some real case scenarios. From this exercise it was possible to identify the following list of relevant features:

- Perfect match (100/100) with must-have requirements
- Perfect match (100/100) with nice-to-have requirements
- Education in relevant field/discipline
- Formal education level
- Relevant professional experiences
- Having professional experiences in a different field of interest
- Quantity of previous experiences (the more the better)
- Duration of previous experiences (the longer the better)
- Relevance ONLY of the latest work experience
- Prestige of past experiences
- Meeting ONLY the hard skills
- Language Level requested in the job offer
- Having the required licenses
- Candidate matching Diversity & Inclusion policies' criteria
- Cultural and Value fit for the company
- Availability to flexible working hours
- Distance from candidate and job's location
- Not being overqualified
- Law and policy compliance (visa requirements etc)





Communication and social skills
 Different interesting hobbies
 Motivation and Resourcefulness
 Lack of experience compensated by a strong educational background

After configuring the tool, participants reflect on the outcomes, discussing the final ranking, decision-support explanations, and potential improvements. The exercise concludes with a group discussion on the tool's perceptions, addressing potential hazards, trustworthiness, and fairness.

BIAS Co-Creation Workshop - The Candidate Ranker

Insert your group name and press enter...

Score and rank the candidates according to the following features:

	Features:	Priorities:
Feature 1:	Select one... ▼	Select priority level... ▼
Feature 2:	Select one... ▼	Select priority level... ▼
Feature 3:	Select one... ▼	Select priority level... ▼
Feature 4:	Select one... ▼	Select priority level... ▼
Feature 5:	Select one... ▼	Select priority level... ▼
Additional custom features:	eg: feature6, feature7, feature8...	eg: must-have, nice-to-have, nice-to-have...
Additional Instructions...		
SUBMIT		

Figure 20 Interface of the Candidate Ranker Tool

11.2.2 The Mitigation Tool: Identifying and mitigating BIAS from CVs and cover letters – Methodology

Overview:

In the deliverable D2.2 the BIAS Consortium made an initial attempt to define diversity biases of AI-driven systems¹⁸. The term 'bias' should be understood neutrally as a deviation from a standard, encompassing statistical, moral, and legal dimensions. The same AI application can be biased based on one standard but not on another. The understanding of bias adopted by the Consortium defines it as systematic and unfair discrimination by computer systems (Friedman & Nissenbaum, 1996). Regardless of the type or origin of bias in AI applications, the BIAS Consortium aligns with existing literature in acknowledging that these applications are created and operated by humans. This acknowledgment implies that human designers and users of AI make subjective value judgments and assumptions during the design process, and these decisions are likely to be influenced by their implicit or explicit biases. In essence, the human factor plays a significant role in shaping the biases present in AI systems. To explore these considerations, the Mitigation Tool was developed as a simulation system to assess people's perceptions around the Debiaser. It aims to observe the role of the human factor in shaping biases in AI systems and how the AI system influences human perceptions of biases.

The mitigation tool is designed to simulate the functionalities of the word-embedding based Debiaser, particularly during the screening phase, supporting the recruiter evaluating candidates. The system helps ensure that candidates are evaluated based on relevant qualifications rather than factors such as gender, ethnicity, or other protected characteristics. This particular simulation tool is very generic, but it would give a good indication of the overall perception of such a system within a diverse user base. The objective of deploying this exercise is to initiate meaningful discussions about bias concepts and varying perspectives on mitigation strategies.

¹⁸ For details please see Deliverable D2-3.2.3





The “Mitigation tool” functionalities

The structure of this tool is very similar to the “Candidate Ranker”, it works as well on ChatGPT APIs, and requires a user input to function. The user can select up to three from a list of bias types or insert a bias type of their choice; the prompt is crafted similarly to the first system, encompassing specific instructions, the candidates’ profiles (including cover letters and resumes), and the job advertisement, which is only partially used by the system in this case, as the biases identified are mainly extracted from the candidate’s data (see [Annex 14](#)). Nonetheless, it is pertinent to include the job advertisement to address potential position-specific biases; for example, a work that is mostly considered for man only. Bias categories cover a wide range of discriminatory factors such as gender, race, disability, and the prompt instructs ChatGPT to focus on these when analyzing each candidate. Additionally, the instructions include the request to provide a small mitigation solution for every bias element identified. The final prompt structure has therefore a similar structure to the candidate ranker, as well as the output format, which is saved as text and displayed on the application.

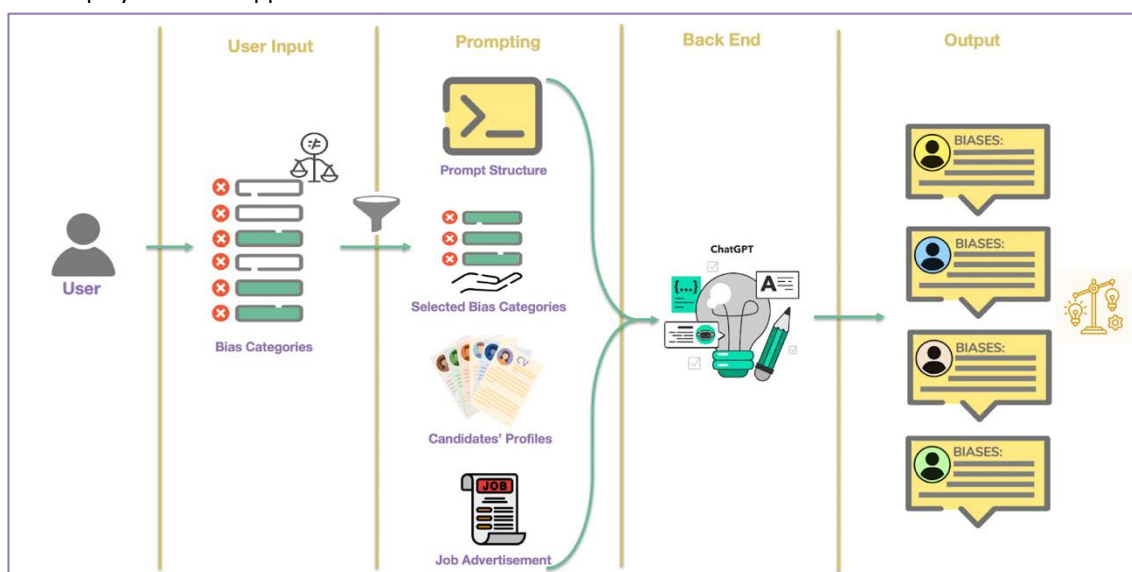


Figure 21 Simplified User Journey of the Mitigation Tool

The Exercise:

This exercise starts with group discussions on potential biases present in candidate profiles, focusing on listed factors which are “Age”, “Gender”, “Ethnicity”, “Gender identity”, “Sexual orientation”, “Diverse abilities”, “Skin color”, “Race”, “Religion”, “Social status” and “Economic background”. Groups select three key biases among the ones listed (or can insert their own bias category) and one additional bias specific to HR practices: onformity bias , horn/halo effect, confirmation bias, illusory correlation, etc¹⁹. Participants then learn how the “mitigation tool” works, prompt the system with the chosen biases, review the system's output, and reflect on the identified biases and mitigation strategies. The group compares biases identified by both the system and themselves, evaluates the reasonableness of the system's explanations, and discusses the tool's strengths and potential improvements. The final session involves a plenary discussion, exploring perceptions of the tool's hazards, trustworthiness, and fairness.

¹⁹ The compilation of diversity biases in the labor market is a condensed summary of the discussions presented in Deliverable D2-3.2.5, highlighting the prevalent biases encountered in the selection process. (Harver, 2022) (Black & van Esch, 2020) (Barragán Díaz et al., 2019)(Equalture, 2023).





Figure 22 Interface of the Mitigation Tool

In [Annex 16](#) there are all the guidelines for the exercise, as well as the technical instructions on how to use the tool.

11.3 Second Block: Trustworthiness – Methodology

11.3.1 Activity Overview

The second session, held after the lunch break, delved into theoretical discussions surrounding trustworthiness. In the second co-creational workshops across the consortium countries, one activity aimed to brainstorm initial technical requirements for the Debiaser's deployment. However, participants found the abstract concept of the tool somewhat challenging to grasp, leading the activity to evolve into a brainstorming session focused on desired functionalities of a Debiaser. This session was strategically scheduled after participants had interacted with a simulation of the systems, providing them with a clearer understanding to analyze the requirements more effectively.

The objective of this exercise is to gather technical requirements from three distinct perspectives: the technical needs of AI, the needs of HR, and the essential needs of workers and workers' representatives. Each expert was tasked with identifying critical aspects of such tools within their field of expertise and contributing to the design of a solution. The structure of the exercise aligns with this logic: based on the seven requirements outlined by ALTAI, participants were divided into groups according to their areas of expertise. They were then prompted to contemplate potential issues, solutions, and actions to address every emerging challenge. The aim of this activity is to structure the needs and requirements of the Debiaser in alignment with the ethical guidelines proposed by AI HLEG, considering three different stakeholders' perspectives. In doing so, it ensures the deployment of a trustworthy and fair system.

11.3.2 The ALTAI Requirements

In the deliverable D.2 it is explained that the BIAS Consortium, when focussing on technical fairness formulations, plans to evaluate the Debiaser's compliance using the trustworthy AI assessment list (ALTAI)²⁰: the High-Level Expert Group on Artificial Intelligence (AI HLEG) views fairness as an ethical

²⁰ <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>





principle with both substantial and procedural aspects, aiming for equal distribution of benefits, non-discrimination, and transparent decision-making in AI applications; with this purpose, it developed a self-evaluation tool to measure trustworthiness in AI-driven systems, the ALTAI. Following a piloting process where over 350 stakeholders participated, an earlier prototype of the list was revised and translated into a tool to support AI developers and users in developing Trustworthy AI²¹.

The tool supports the actionability of the key requirements outlined by the Ethics Guidelines for Trustworthy Artificial Intelligence, presented to the European Commission, in April 2019. The Ethics Guidelines introduced the concept of Trustworthy AI, based on seven key requirements:

1. Human Agency and Oversight

It emphasizes the importance of AI systems supporting human agency, respecting autonomy, and being subject to effective human oversight to ensure ethical and trustworthy AI development and deployment.

2. Technical Robustness and Safety

It emphasizes the importance of technical robustness and safety in AI systems, including resilience to attacks, general safety measures, accuracy considerations, and strategies for reliability, fall-back plans, and reproducibility. This ensures that AI systems are developed with a preventative approach to risks, behave reliably, and minimize harm, aligning with the goal of achieving Trustworthy AI.

3. Privacy and Data Governance

It emphasizes the importance of safeguarding privacy and implementing robust data governance measures in AI systems. This includes assessing and addressing the impact on fundamental rights, considering data protection implications, and aligning with relevant standards to ensure responsible and ethical AI development and deployment.

4. Transparency

It underscores the importance of transparency in AI systems, covering traceability, explainability, and effective communication of capabilities and limitations. This transparency is essential for building trust, especially in cases where full explainability is challenging, ensuring users are informed and able to contest decisions when necessary.

5. Diversity, Non-discrimination, and Fairness

It underscores the importance of diversity, non-discrimination, and fairness in the development and deployment of AI systems. It emphasizes strategies to prevent bias, clear definitions of fairness, universal design principles for accessibility, and active stakeholder participation to ensure ethical and equitable AI.

6. Societal and Environmental Well-being

It emphasizes the ethical and societal considerations associated with AI systems, including their impact on well-being, the environment, work, skills, and democracy. It encourages responsible development and deployment practices to ensure positive societal outcomes.

7. Accountability

It underscores the importance of creating accountable AI systems by implementing auditability, effective risk management, external oversight, continuous monitoring, and mechanisms for reporting and redress. It aims to ensure responsible and transparent AI practices throughout the system's life cycle.

²¹ For more details please consult the Deliverable D2-3.2.7





11.3.3 The Exercise in practice

This exercise focuses on evaluating the application of a bias-free AI system, based on the experiences of the first block of activities and the interactions with the simulation tools. Participants are asked to consider specific requirements and respond to various aspects, in a brainstorming setting:

- Identify Critical Issues (Red Post-its):
Participants are prompted to pinpoint potential problems or critical issues related to the Debiaser for each listed requirement. These concerns are recorded on red post-its.
- Outline System Solutions (Green Post-its):
For each requirement, participants are tasked with proposing solutions that can be implemented during the development of the Debiaser to ensure compliance. These solutions are documented on green post-its.
- Identify Recruiters' Actions (Blue Post-its):
Participants are expected to identify actions that recruiters and other stakeholders can take to ensure compliance with the Debiaser for each requirement. These actions are recorded on blue post-its.
- Develop Metrics (Violet Post-its):
Participants are asked to create measurable metrics for assessing the fulfillment of each requirement by the Debiaser. These metrics are documented on violet post-its.

The use of color-coding (red for issues, green for system solutions, blue for recruiter actions, and violet for metrics) aids in organizing and visually representing the responses. This exercise encourages thoughtful consideration of potential challenges, solutions, actionable steps, and measurable criteria in implementing a bias-free AI system for recruitment decisions. Each table is equipped with a poster, colored post-its, and pens. The brainstorming task engages participants in generating and organizing ideas within each group. The poster ([Annex 17](#)) displays the 7 ALTAI key requirements along with their explanations. Utilizing a problem-solving approach, participants are prompted to consider potential applications for each key requirement. For example, they may contemplate the hazards the Debiaser could face concerning 'Human Agency and Oversight' (red post-it). Subsequently, they are encouraged to propose solutions for these hazards (green post-its), identify necessary actions and responsible parties (blue post-its), and suggest metrics for evaluating the solutions (violet post-its).

11.4 Third Block: Learning Needs – Brainstorming

During the international workshop's third and last activity, participants engaged in a collaborative brainstorming session. The aim was to gather ideas and insights from the various groups, contributing to the development of tailored training materials and addressing the learning needs of each category of stakeholder. The facilitator guided the groups through a series of questions, with the rapporteur summarizing the key points that surfaced during the discussions (see [Annex 18](#)). This interactive session provided a platform for diverse perspectives to be shared, fostering the generation of comprehensive training strategies to meet the needs of different stakeholders.

1. Have you ever attended, or are you aware of any course or training (both within and outside university contexts) dealing with the topic of bias in AI and in particular in AI used in recruitment and selection processes?
2. If you have to attend a capacity building on the topic of bias in AI systems and in particular in AI systems to be used in recruitment and selection, which knowledge/information would you expect to get?
3. Which skills and competencies would you expect to develop/improve?





4. Do you have any case study to share (both from your organization and elsewhere) about the use of AI technologies that lead to bias? In particular, any on the use of AI technologies used in recruitment and selection that ended up favoring a particular category of candidates over others? With case studies we mean any real case/example containing either positive or negative lessons learnt that could represent learning experiences for participants to the BIAS capacity building programme.
5. According to your workload and working schedules which should be in your opinion the ideal length of a capacity building course? Would be one day and an half ok? Would you preferred having a more compact session or multiple sessions spread in one or more weeks
6. Would you prefer an in person or online format?

To be noted how results from this specific activity are not reported in this Deliverable as the analysis of the findings will feed into D5.1 instead.





12. The International Workshop Co-Creation - Results

12.1 First Block: simulating the use of AI based systems in Recruitment and Selection - Findings

As discussed in Chapter 8, the initial set of activities took place on the morning of December 7, 2023. Participants were divided into six groups, each consisting of seven to eight individuals. To ensure a diverse composition within each group, at least one HR officer and one AI specialist were included. The groups were thoughtfully named after fruits (APRICOT, BANANA, CHERRY, LEMON, MANGO, and PEACH) for easy data tracking.

The groups were stationed around tables in the main room, with one exception – a group was situated in the bar room to accommodate the needs of a participant with hearing impairment. Carefully printed materials required for the activities were provided to each group. Additionally, group facilitators utilized their laptops to illustrate activities, while rapporteurs used theirs for note-taking.

The first block covered the whole morning and comprised a series of activities, featuring two exercises involving simulation tools and concluding with a plenary session. In sections 9.2 and 9.3, the results of these activities are presented with a consistent structure:

- 1. Exercise execution:**

An initial explanation of how the exercise was conducted, with a focus on the quantitative data derived from the activity.

- 2. Strengths and Shortcomings of the tool:**

A subsequent paragraph detailing the reflections on the experience highlighted by participants, structured as strengths and weaknesses of the tool, with the formulation of various recommendations stemming from the shortcomings identified.

- 3. Recommendations to the Debiasser developers:**

For the first simulation of the Candidate Ranker there is an additional dedicated paragraph focusing on recommendations directed specifically to the developers of the Debiasser tool. The workshop provided an opportunity to delve deeper into the technical aspects of the tools, leading to the formulation of specific requirements for the Debiasser developers.

12.2 Activity One: The Candidate Ranker – Results

12.2.1 Exercise execution

During the first activity of the morning block, participants had the opportunity to interact with the candidate ranker: each group had the freedom to engage with the tool as many times as desired. Most teams embraced the recommended “try and see” approach, where they conducted initial iterations, adjusted inputs, and observed subsequent changes in the output. The following indicates the number of iterations conducted by each group, with an average of three iteration per team:

The charts provided below illustrate the frequency of each requirement (must-have and nice-to-have) chosen during the exercise across all iterations. Examining the pre-set features reveals intriguing insights into their perceived relevance. Notably, 15 out of 24 features garnered consideration at least once when establishing must-have requirements, while only 9 out of 24 were factored into considerations for nice-to-have qualities. This initial emphasis on essential requirements suggests a primary focus on streamlining candidate selection, with more dynamic changes occurring in must-have features during various





interactions, whereas nice-to-have requirements remained relatively stable and were less frequently modified or discussed.

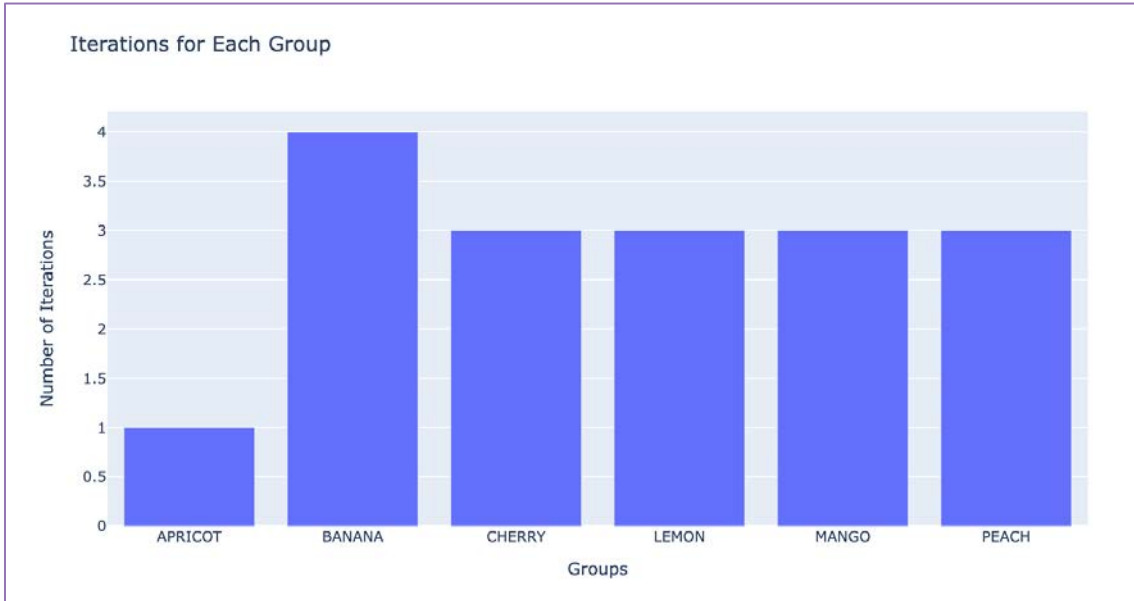


Figure 23 Iterations for each group with the Candidate Ranker

"Relevant professional experience" emerged as the top must-have requirement, with a specific emphasis on the importance and duration of recent experiences: indeed, the "Duration of previous experience" also ranked as the most selected nice-to-have feature. Participants highlighted the subjective nature of evaluating the relevance of professional experiences, advocating for HR professionals involved in hiring to precisely define this criterion. The perception of "Motivation and Resourcefulness" varied among participants, being among the most selected must-have requirements in some groups but falling into the nice-to-have category in others. Participants acknowledged the challenge of accurately gauging these attributes solely from resumes and cover letters. Similarly, the difficulty in assessing "Communication and Social Skills" through written documents was noted, highlighting the complexities involved in evaluating certain qualities during the hiring process.

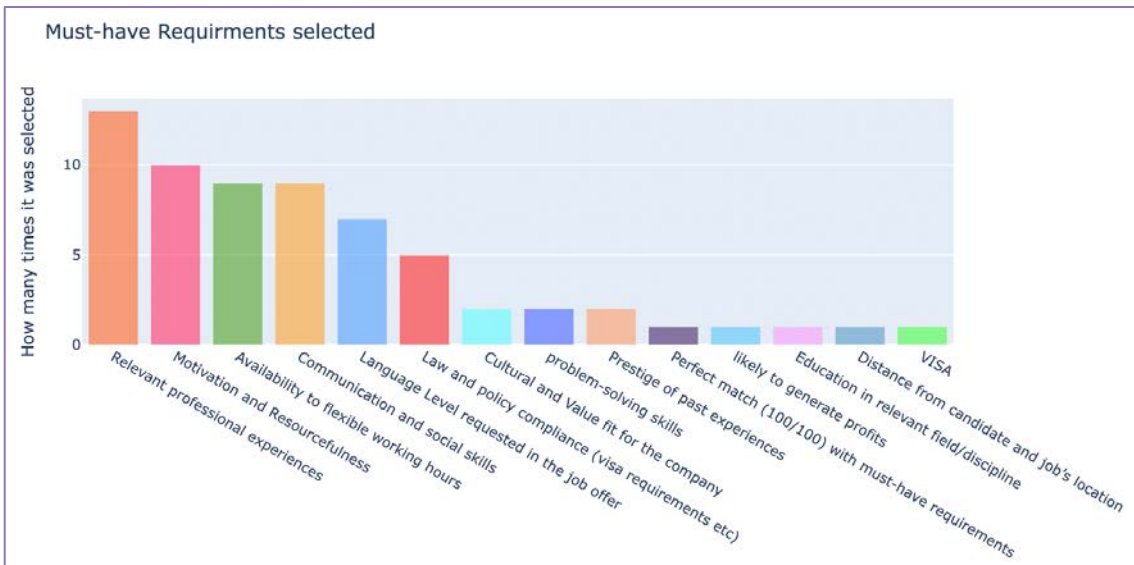


Figure 24 Must-have Requirements selected in the Candidate Ranker



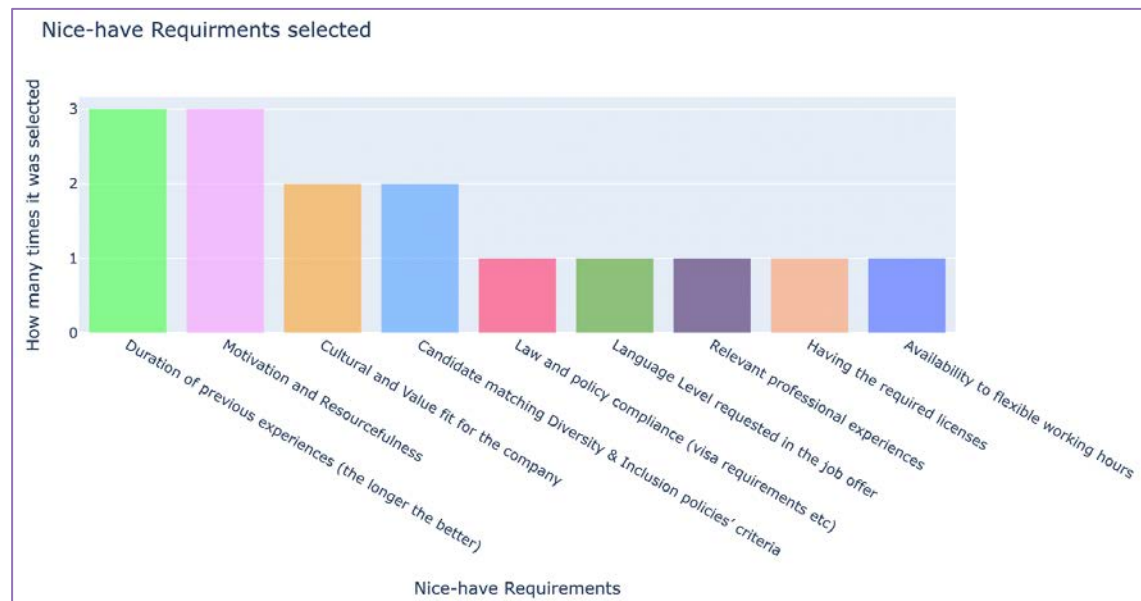


Figure 25 Nice-to-have Requirements selected in the Candidate Ranker

There was a collective discussion on how the various features should be prioritized, such as relevant professional experiences, language proficiency, flexibility in working hours, and compliance with laws and policies. Discussions also centered around the importance and ambiguity of "Cultural and Value Fit", recognized as one of the most crucial requirements. However, participants acknowledged the challenge of assessing this aspect at the current stage of the hiring process, emphasizing the absence of effective means for its evaluation. The issue of cultural fit and work-life balance was discussed with mixed views, reflecting the subjective interpretation and/or different possible framing of these concepts. Additionally, participants emphasized the relevance of "Law and Policy Compliance" and discussed potential biases, for example towards a university dropout, or based on socio-economic situations or gender. Concerning language requirements, several groups opted to introduce their own feature, expressing a belief in the necessity of providing the system with more robust input.

To gain a clearer understanding of where people tend to focus during candidate screening, the selected features can be divided into the following macro themes:

- **Job relatedness:** features that evaluate the pertinent characteristics of candidates based on the job offer. These include also all the hard skills and elements that are mentioned in the job offer as requirements: 'Relevant professional experiences', 'Availability to flexible working hours', 'Language Level requested in the job offer', 'Perfect match (100/100) with must-have requirements', 'Education in relevant field/discipline', 'Having the required licenses', 'Distance from candidate and job's location'.
- **Law and policy regulation:** all features associated with law enforcement and adherence to company policies: 'Law and policy compliance (visa requirements etc)', 'VISA', 'Candidate matching Diversity & Inclusion policies' criteria', 'Law and policy compliance (visa requirements etc)'.
- **Soft skills:** all features related to the assessment of candidates' interpersonal and personal attributes: 'Motivation and Resourcefulness', 'Communication and social skills', 'problem-solving skills'.
- **Cultural fit:** the feature that evaluates how well a candidate may integrate with the company's work ethic, philosophy, and values: 'Cultural and Value fit for the company'.





- **Positive impression:** include all features that may not be directly relevant to the job but contribute to enhancing the overall appeal of a candidate's CV. These features are elements that, while not essential to job performance, can positively influence how a candidate is perceived during the evaluation process: *'Prestige of past experiences', 'likely to generate profits', 'Duration of previous experiences (the longer the better)'*.

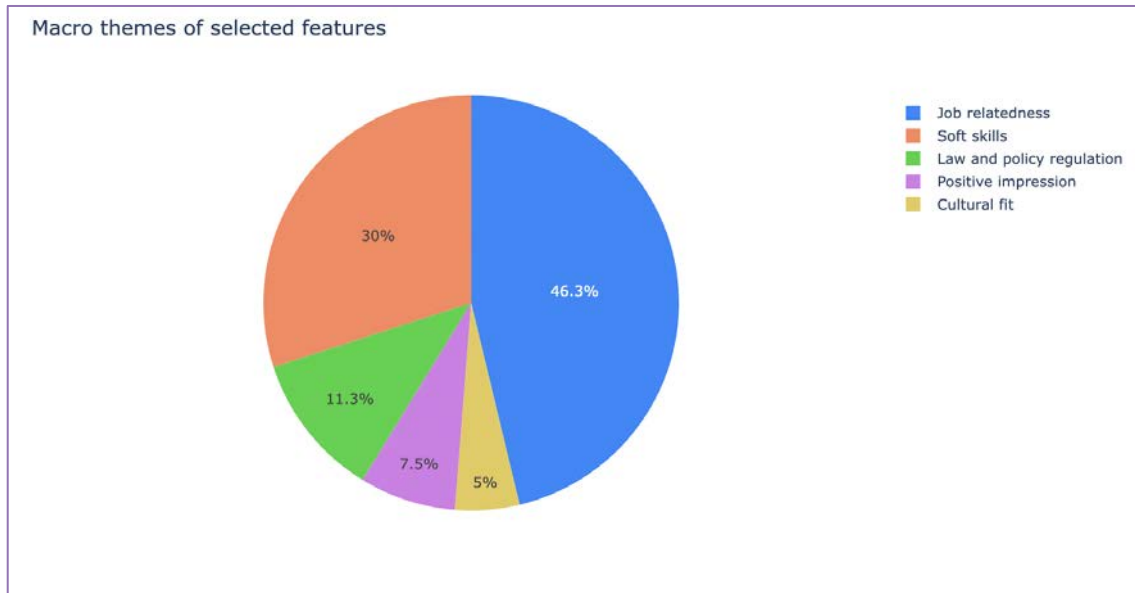


Figure 26 Macro-themes of the selected features: distribution

Almost half of the selected features fall under the category of job relatedness, signifying its predominant importance in candidate evaluation. Soft skills make up around 30% of the features, emphasizing the recognition of their significance in assessing candidates beyond technical qualifications. Approximately 10% of the features pertain to law and policy compliance, acknowledging the fundamental role of law in filtering out candidates but also noting that it's often taken for granted in the initial evaluation, hence the low selection.

Positive impression and cultural fit collectively cover 12.5% of the selected features. While cultural fit may be considered highly important, participants noted the difficulty of accurately assessing it based solely on cover letters and CVs. As a result, cultural fit, along with positive impressions, is seen as elements to be considered more extensively in later stages of the evaluation process, particularly during interviews. The distribution of emphasis on these different elements reflects the participants' nuanced approach to candidate selection, taking into account practical considerations and the limitations of early-stage evaluations.

After the selection of the features to prioritize, participants observed the outcome of the Candidate Ranker. The machine distribution of ranking positions for each candidate is illustrated below: Anna predominantly occupies the first position, depicted by a Gaussian centered around Ranking=1 (x) with minimal variance (a tall and narrow curve). Conversely, Mohammed and Mark are frequently positioned in the second or third places, while the last three candidates exhibit greater variability in their rankings. This pattern suggests that the tool finds it relatively straightforward to identify the "best" candidate, but faces challenges in assessing the relative merits of all candidates, particularly when making direct comparisons among them.



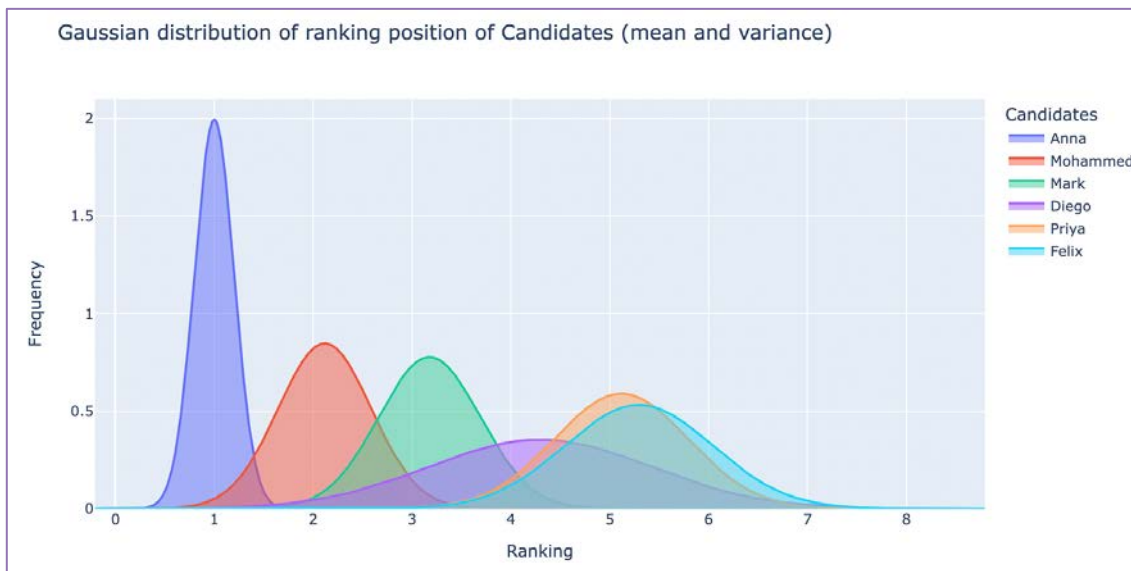


Figure 27 Gaussian Distribution of the ranking positions calculated by the Candidate Ranker

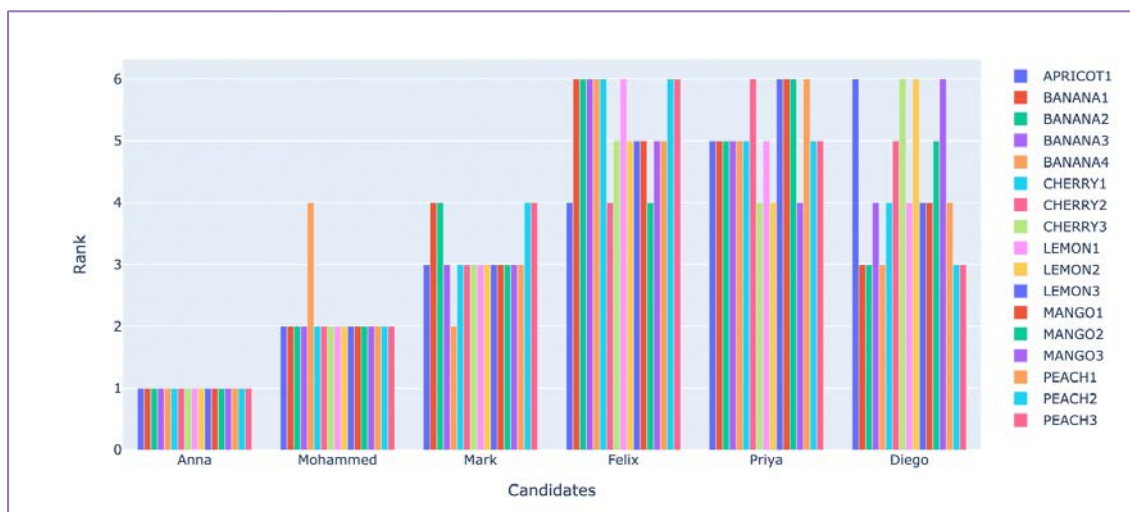


Figure 28 Detailed distribution of the ranking positions for every group and its iterations

12.2.2 Strengths and shortcomings of the tool

During the Candidate Ranker exercise, various aspects of the tool were highlighted and thoroughly discussed. The overall perception of the tool was positive, marking a significant moment where participants could actively engage and experience first-hand the practical applications of AI in this domain. The tool was deemed highly useful in facilitating a more detailed and effective analysis of the Debiaser requirements, risks, and potential solutions. This hands-on experience contributed to a deeper understanding of AI's role in the field of recruiting and provided valuable insights for further development and refinement of such tools. That being said, during the exercise, many criticalities were identified and discussed. The mixed reactions from the workshop's participants reveal that while AI-driven tools like the "Candidate Ranker" can be helpful for HR tasks, they need to be carefully calibrated and their logic made transparent to avoid potential biases and meet user expectations. The simulation was recognized as a good starting point but improvements were necessary for a more holistic, fair, and beneficial candidate ranking.

**Prioritizing the features:**

While it was interesting to explore the relevant factors in screening candidates, it was highlighted that among must-haves, not all are equally important. The variance in opinions emphasized the importance of weighting these features to reflect different levels of importance. Participants also expressed the desire for a breakdown of how each priority or feature contributes to the final ranking and more transparency in the logic mechanism. As such, the tool should highlight elements that contribute to the scoring directly in the text, in that way, recruiters could have more visibility over the most relevant information.

Interpreting certain features posed a challenge during the exercise. For instance, the criterion "flexibility" raised concerns among participants, with some suggesting that the term might be a euphemism for anticipating overtime. This highlights a significant issue in the recruiting process, where the interpretation of terms can vary subjectively among individuals. It's worth noting that this challenge is more closely tied to the human recruiter making decisions rather than the AI supporting the process.

The tool's ranking:

When examining the tool's final candidate ranking, there was a mix of agreement and disagreement among participants. Some found it surprising, while others did not understand why their ideal candidate was not placed higher. When observing all the tool's ranking lists, the consistent outcomes of the machine in every iteration underscores two key observations. Firstly, it indicates a level of reliability and consistency in the outcomes produced by ChatGPT. This consistency contributes positively to the perception of the system, demonstrating that it doesn't act randomly. This characteristic helps avoid portraying the AI as a black box, instilling a sense of predictability and understanding in its behavior, which can be crucial for user trust and transparency. However, it became apparent that participants had limited influence on the tool's prioritization of features. Regardless of the selected features, the resulting ranking exhibited minimal changes. This observation raised concerns about the potential risks associated with a tool that operates with a high degree of independence from human control. Participants recognized this as a potential drawback and emphasized the importance of implementing a human-in-control approach for AI systems. This experience provided valuable insights into the balance needed between automation and human oversight, highlighting the necessity of ensuring that AI tools remain subject to meaningful human input and guidance. Participants expressed a desire for more transparency from the tool, especially in terms of how it calculates the ranking.

Identification of relevant information:

Generally, participants gravitated towards similar concerns, resulting in comparable outcomes across all groups. Although participants were not explicitly instructed to formulate their ranking lists, the prevailing perception was that each candidate exhibited strengths and weaknesses open for further discussion. However, certain aspects were identified as potential filters that the machine failed to recognize. For instance, Anna lacked proficiency in the required language, a factor that was intuitively evident to participants, making her a less suitable candidate. Despite explicit prompts, the system consistently ranked Anna in the top position, overlooking this crucial factor. People also noted that the tool did not seem to catch errors and discrepancies in the job advertisement, failing to pick up errors or recognizing variations in how information was presented.

Bias in the decision:

During the course of the workshop, participants evaluated candidates with various diverse backgrounds and situations. For instance, some candidates had visa issues, unique personal situations (like being a single parent), or were from unconventional backgrounds like being a food influencer on Instagram. Participants noted the potential for bias of recruiters in these situations and pointed out that in reality, these factors could be nuanced and discussed during an interview process. When participants tried changing the configuration settings, they noted that the ranking did not change much. In some cases, this





led to a mistrust in the functionality of the tool, especially when important must-have factors seemed to be ignored. These discrepancies led to a sense of bias within the tool, particularly when the tool still ranked a candidate highly despite incontrovertible issues, like visa problems. Several participants highlighted the issue of potential bias impacting the ranking: personal lifestyle choices or family circumstances sometimes seemed to be weighed more heavily than necessary. In the plenary session, participants also recognized the hazard of system-induced biases, which could significantly influence HR personnel's perceptions. However, they acknowledged that such a tool could support the hiring process, especially for roles with a large number of candidates.

Clarity of the explanations:

When asked if the provided explanations were sufficient for decision-making, the participants' responses were varied. Some criticized the tool's lack of clear explanations on its ranking process and how it calculated scores, necessitating reverse engineering attempts to understand the tool's logic. Participants acknowledged that the tool was good at summarizing and identifying several aspects from the resumes and cover letters. However, they felt there were functionalities missing from the tool, like the ability to provide feedback to candidates to improve future applications.

Fairness of the process:

In the plenary session, some interesting points were highlighted. Concerns ranged from the tool drawing incorrect conclusions to potentially manipulating candidates' information to make them appear better than they actually are, and they stressed the tool could hallucinate data (as it actually did in some cases). For these reasons, it was emphasized that human interaction should not be totally eliminated from the recruitment process, with participants advocating it to be used strictly as a supportive rather than decisive element in the selection process. Participants felt apprehensive about finding the tool fair: trust issues were exacerbated by the perception that the tool could be biased, possibly due to gender or other factors not directly related to candidates' qualifications. They also noticed that the tool's utility hinged on receiving good input data, which is not always the case in this field. Several overarching improvements were suggested to increase the tool's trustworthiness: anonymization of details that may lead to bias, implementing a function to suggest follow-up questions, and creating a domain-specific language model for recruitment.

It was reiterated that while the tool isn't expected to be inherently fair, it should be exact and precise in the data it processes. The ultimate interpretation, and the fairness of the decision, should be left to the HR professionals.

Trustworthiness of the process:

The trustworthiness of the tool was closely tied to its perceived fairness. Participants did not necessarily find the tool to be fair. They did, however, acknowledge its usefulness, particularly in dealing with a large number of CVs. It was observed that the tool accurately simulated the HR process but the outcomes differed when compared with human decision-making. Generally, recruiters often grapple with uncertainties when assessing candidates solely based on cover letters for various reasons. Some participants pointed out that candidates with poor writing skills might be at a disadvantage. Simultaneously, others observed that a significant number of individuals utilize AI-generated text to enhance their presentation. As a result, there was a prevailing belief that standardized inputs, rather than personal cover letters, would contribute to a more level playing field in the evaluation process.

The role of social justice:

Participants concluded by examining the social justice factor in hiring practices: it was suggested that the measure of fairness should consider company objectives and societal context. For example, a seemingly unbalanced gender distribution might be a result of inherent demographics in that field, so adjustments should be company/sector-specific and take into account broader recruitment policies.





Precision of the system:

The tool was deemed unreliable in terms of extracting precise information. Fairness testing was recommended, proposing controlled trials involving tweaking variables to study changes in the outcomes. Recommendations for improving the fairness and reliability of the tool included a suggestion to abandon the ranking system in favor of adopting a matching features system. Participants proposed a system solely focused on identifying relevant elements in a candidate's profile, favoring a matching approach over a decision-based system. They also advocated for enhancing the precision of the tool's summary output.

Furthermore, AI experts emphasized the importance of prioritizing recall over accuracy. In machine learning, accuracy measures a model's ability to identify relevant data points, while recall assesses its capability to identify all correct predictions within a dataset. While high accuracy is crucial for recruiters making selections, for a supporting tool like the Debiaser, ensuring the consideration of all potential right-fit candidates is paramount. Even if this results in some less suitable candidates making it into the shortlist due to less accuracy, the emphasis on inclusivity is vital for meeting participants' essential requirements.

12.2.3 Recommendations to Debiaser developers

The Candidate Ranker simulation facilitated a detailed exploration of the technical aspects of the tools, resulting in the formulation of specific requirements tailored to enhance the functionality and effectiveness of the Debiaser. Participants were tasked with selecting requirements and evaluating their alignment with candidates' skills and capacities; from an algorithmic standpoint, key insights include the following:

- **Macro-Features and hierarchical sub-categories**

Participants emphasized the significance of a well-structured approach to requirements. While a list of features might be beneficial in a simulation setting, it is deemed insufficient for a comprehensive candidate assessment. It was noted that each requirement comprises multiple evaluation elements, suggesting the ideal construction would involve macro-features encompassing various sub-groups in a hierarchical structure. For instance, "Educational Background" could include sub-categories such as "Educational Level," "Validity of the Institution," "Field Relevance," "Additional Metrics," and so on. Moreover, it is deemed essential to establish criteria for assessing the "relevance" of the study, leading to the inclusion of features like "Courses" and "Thesis topic."

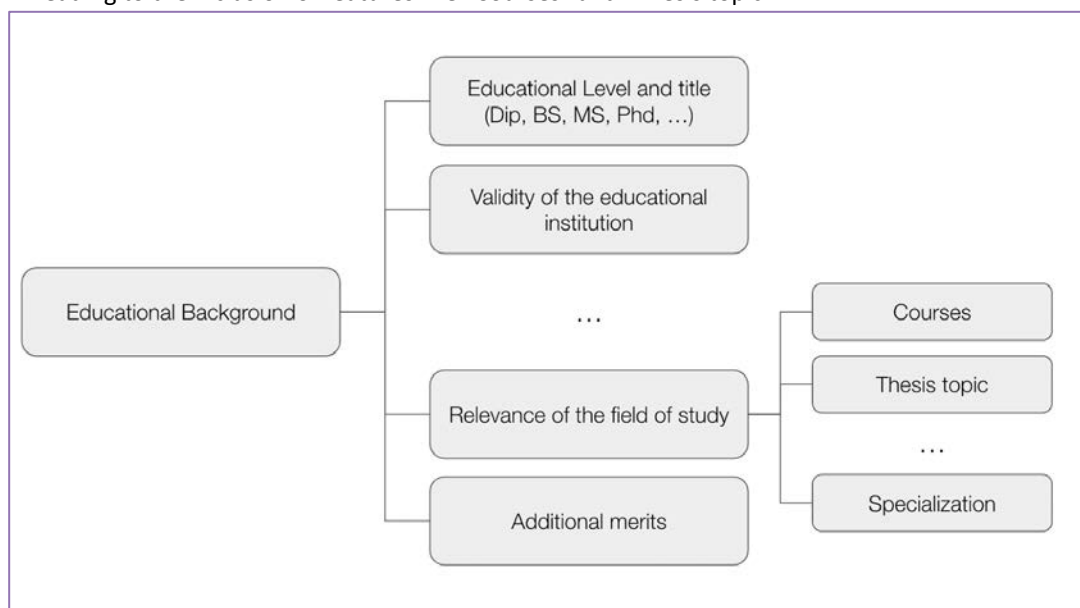


Figure 29 Example of the hierarchical decomposition of every feature in sub-categories





- **Weighted Prioritization**

Participants observed a limitation concerning the designated level of importance. In the simulation system, the available options for importance levels were restricted to just two choices (must-have or nice-to-have). However, participants expressed the view that this binary classification was insufficient for assigning accurate weights to each requirement. To address this concern, they suggested that the decision support system should incorporate a more comprehensive scale, allowing for nuanced importance configurations, such as weights ranging from 1 to 10.

- **Clusters of combined features**

One important discussion concerned how to include in the decision the variety of acceptable profiles and personas for a given job offer: different profiles can have different sets of skills that turn all of them into valid candidates. So instead of thinking about categories and subcategories of requirements, the decision support system should consider a variety of clusters of different skills that are valid for a position; to give an example, a job ad for a marketing specialist might require a degree in a related field, but candidates with no degree could still be considered if they have strong enough professional experiences. The single requirement is not necessarily relevant for the final evaluation, what matters is the combination of all the skills; it is the combination of the fulfillment of multiple requirements that is relevant to assess a candidate, every candidate can have a different combination of such requirements and still be a good fit for the company. The ideal would be to include all the combinations that are acceptable for a job role, each one with a certain threshold to reach. The thought that came out was to build the evaluation features and fulfillment as a decision tree. In this tree, if a candidate meets a basic requirement, they would then be assessed against subsequent criteria. An essential discussion revolved around incorporating the diversity of acceptable profiles and personas for a given job offer. Different profiles may possess distinct sets of skills, rendering them equally valid candidates. Instead of framing the decision-making process around categories and subcategories of requirements, the decision support system should embrace a range of skill clusters deemed acceptable for a position. For example, a marketing specialist job advertisement might traditionally require a related degree, but candidates without a degree could still be considered based on robust professional experiences. The emphasis lies not solely on individual requirements but on the amalgamation of skills. The decisive factor is the combination of multiple requirements fulfilled by a candidate, allowing for unique combinations that still align with the company's needs. The envisioned ideal involves encompassing all acceptable combinations for a job role, each with a defined threshold. This concept suggests a departure from linear weighting, where each feature is assigned a weight based on its individual importance. Instead, the proposal involves structuring features sequentially, adjusting the weights of each subsequent feature based on the values derived from the previous evaluation. In our example, if the macro-feature related to "educational background" attains a high score, the macro-feature of "previous experiences" might carry a lighter weight in the overall evaluation, and vice versa.

- **Multi-criteria fairness metric**

The technical perspective on fairness was affirmed to be domain-dependent, highlighting that various cases can be deemed equally fair despite their differences. This indicates the need, in the assessment of fairness, to consider a range of variables measuring different aspects of fairness. Evaluating fairness should be approached as a composition of these variables, categorizing each element as adequately fair concerning the benchmark.





12.3 The Mitigation Tool: Identifying and mitigating BIAS from CVs and cover letters – Findings

12.3.1 Exercise execution

The second morning activity mirrored the structure and organization of the initial one. Participants engaged in interacting with the Mitigation Tool, a simulation system, and were prompted to provide feedback on their experience. The group composition remained consistent with the previous exercise, and similar to before, participants had the freedom to choose how many times they altered the input provided to the tool. The majority opted to revisit their assigned tasks once or twice.

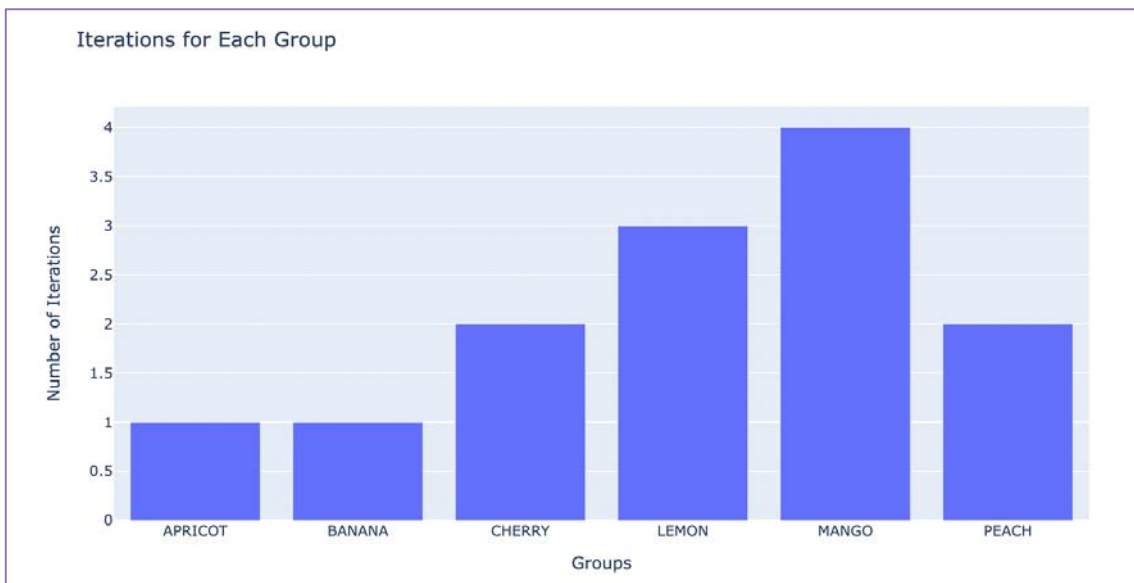


Figure 30 Iterations of each group with the Mitigation Tool

Identifying potential discriminatory elements in the candidates' profiles proved to be a swift task, given the prior discussions in the preceding exercise. Most participants successfully pinpointed the most obvious elements. However, intriguingly, there were instances where participants overlooked factors that the machine detected, and conversely, there were factors noticed by participants that the machine did not identify.

Participants expressed a distinction between "Marital status" and someone's social status, deeming the former a crucial but previously overlooked bias factor. Consequently, they collectively decided to include it. The following graph shows the key biases most frequently selected:



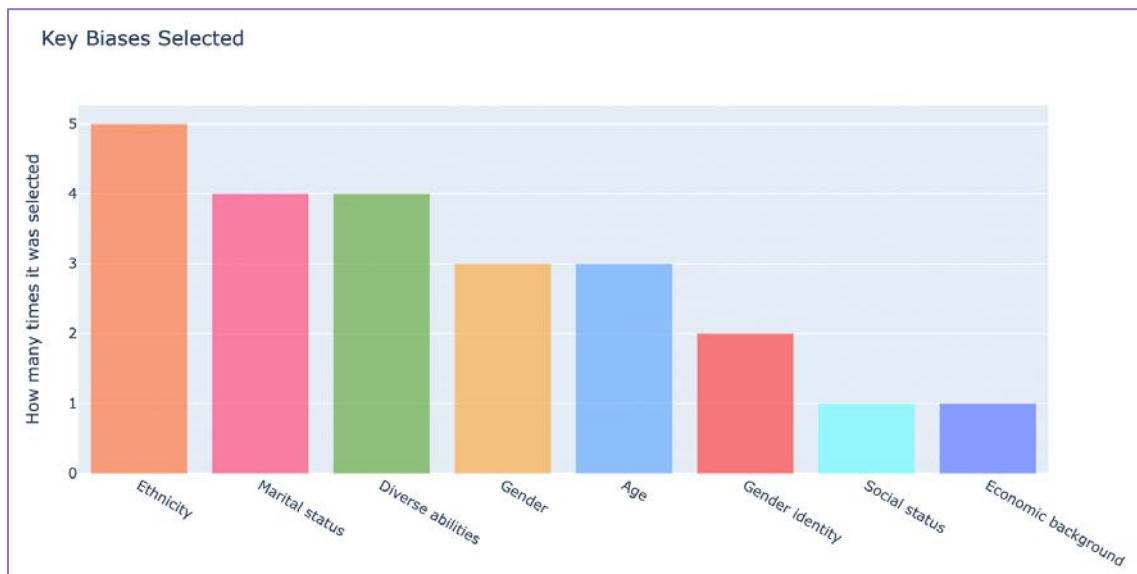


Figure 31 Frequency of the key biases selected by participants in the Mitigation Tool

Participants tended to select key bias elements based on their awareness of their frequency in the data. However, they also speculated about what the system is more likely to identify, revealing an interesting point about how the use of AI can influence people's logical reasoning.

The most frequently chosen key bias was "Ethnicity," followed by "Marital status" and "Diverse abilities." Gender and age were also commonly selected. Notably, when given options related to gender identity ("Gender," "Gender Identity," "Sex"), participants predominantly chose "Gender," followed by "Gender Identity," with none selecting "Sex." While "sex" typically refers to biological attributes distinguishing male from female, "gender" encompasses societal roles and expectations. "Gender identity" represents an individual's internal sense of their own gender, which may not align with the sex assigned at birth. Interestingly, the most considered key bias was related to "Gender", encompassing discrimination towards females in a patriarchy and all forms of discrimination against non-binary individuals. The preference for "Gender" over "Gender Identity" may suggest that participants viewed the former as addressing discrimination towards females and non-binary individuals, while the latter might have been perceived as predominantly related to discrimination against genderqueer, agender, genderfluid, bigender individuals, and less connected to existing discriminations towards self-identified females.

Additional cross-cutting types of bias were discussed. Among them, only two were finally selected for analysis; participants were particularly interested on the type of biases that a tool like the previous one (the Candidate Ranker) would be subject to, as some of them had identified biases in the previous exercise and were curious to see if the machine was "self-aware". One group selected the "Confirmation Bias", which is a cross cutting type of discrimination wherein individuals tend to interpret and remember information in a way that confirms their pre-existing beliefs or hypotheses.



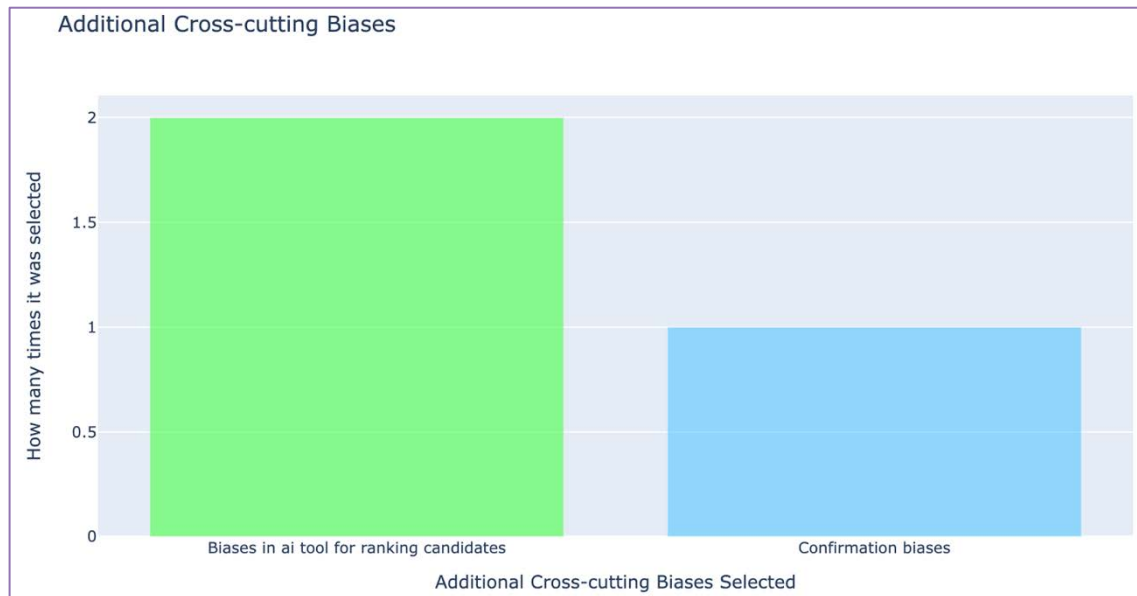


Figure 32 Additional Cross-cutting biases selected by participants in the Mitigation Tool

On average, the system pinpointed 1 or 2 biases per candidate. The most commonly identified biases were associated with disabilities and marital status, predominantly due to participants' selections. However, on occasion, the system proactively identified biases that were not explicitly requested by the participants.

12.3.2 Strengths and shortcomings of the tool

Participants provided varied feedback on the Mitigation Tool, echoing sentiments expressed about the previous system. Similar to the Ranker, the simulation of bias identification was deemed highly useful in comprehending the functionalities of the Debiaser. It allowed participants to form a tangible understanding of the tool's potentialities and risks.

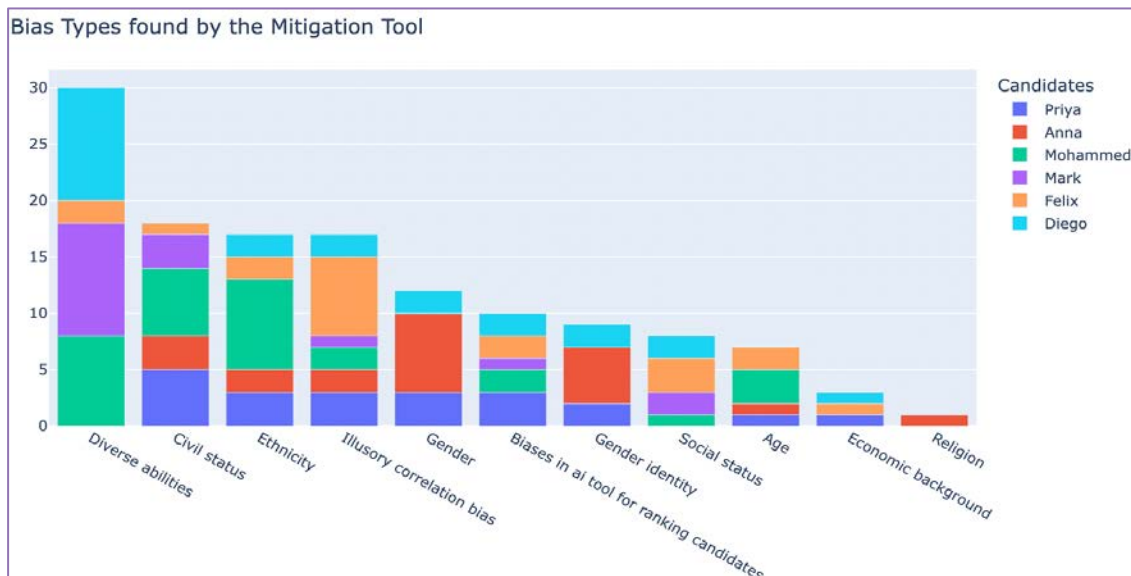


Figure 33 Bias types detected by the Mitigation Tool in the candidates' profiles

Unlike the Ranker, the Mitigation Tool did not involve decision-making, contrasting with the process where the Ranker had to determine the best candidates. This absence of decision-making seemed to elicit a less critical and judgmental response from participants. In general, there was a perceptible openness





among participants to accept this type of technology, possibly due to the non-decisional nature of the tool's function in mitigating biases.

Perceptions over the bias categories:

In general, the participants were satisfied with the list of key biases; as mentioned earlier, certain groups opted to introduce the category of "marital status" to more comprehensively address all discriminations related to the candidate's civil status. Understanding key biases may be relatively straightforward, but cross-cutting types tend to be more domain-specific. As a result, only HR experts found it easier to comprehend these nuances, while other participants, lacking expertise in human resources, faced challenges in grasping the meaning and implications of these cross-cutting types.

The definition of bias:

Participants acknowledged the challenge of selecting and ranking these categories, recognizing the complexity and context-dependent nature of biases. The difficulty arose from the intricate and situation-specific aspects inherent in identifying and prioritizing these biases within the tool. It's worth noting that bias, as a concept, is inherently delicate and complex to identify. Participants exhibited a degree of tolerance toward varying definitions within different categories, recognizing that addressing bias is not an exact science and involves a certain level of subjectivity and interpretation. These considerations align with the observations pointed out by the BIAS Consortium in D2 (3.2.2-3.2.3).

In the preceding exercise, certain participants raised concerns about potential biases in the ranking and explanations, particularly related to gender and marital status. They expressed curiosity to see if the second tool, having the same underlying "brain" as the previous one, would be effective in identifying and mitigating these aspects. This curiosity stemmed from a desire to assess the tool's ability to address and alleviate biases in these specific categories.

Some participants suggested that the inclusion of photos should be avoided to prevent gender and race-related biases, while others felt that an applicant's hobbies, nationalities, and religion played a significant role in selection biases. There was an observation that unconscious biases could be held concerning a candidate's appearance, age, and gender identity. There was also a strong suggestion to include social linguists and psychologists in the process of evaluating potential biases.

The tool's identification of biases:

When using the system, certain biases initially overlooked by the group were pointed out, such as being single or having mental health concerns related to being transgender. However, there were instances where the system either did not detect certain biases or made factual errors, such as the illusory correlation bias associated with night-life activities or misidentifying nationalities, or also potential biases against hobbies. Elements like disability, ethnicity, and language proficiency were inconsistently identified as biases by the tool.

Some of the participants highlighted some tool's problematic assumptions, particularly regarding mental health, transgender status, and marital status: they flagged the tool's assumption that single people might have more time available for work, arguing that more attention should be paid to work-life balance, regardless of marital or family status.

Clarity of the explanations:

Participants recognized some of the tool's strengths, such as its conciseness and clarity. The tool effectively explained why certain elements were considered discrimination factors, providing clear insights even for biases that the groups hadn't initially identified. This transparency facilitated easy evaluation by participants. However, there was also a recognition that there is significant room for improvement. For instance, in some cases, the tool exhibited hallucination by attributing facts of a candidate to a different one or identifying biases in categories that were not instructed. This reaffirmed the requirement for a transparent and clear AI system, under the control of human decisions, as





previously pointed out for the Candidate Ranker. The instances of hallucination emphasized the importance of refining and ensuring the accuracy of the tool to maintain trustworthiness and effectiveness.

Suggestions for enhancement included incorporating a more extensive explanation of the output. This could involve referencing specific parts of the CVs or highlighting text segments that might contribute to bias, thereby improving the individual candidate assessment. The desire for a more detailed explanation aimed at specific CV elements was a common theme in the feedback.

The mitigation phase:

Divergent opinions emerged regarding when in the recruitment process bias should be mitigated – whether before or after ranking. During the initial block of activities, participants envisioned the integration of the two tools in a sequential manner, first ranking candidates and then identifying and mitigating biases. The exact sequence of integrating these functionalities in the Debiaser was never specified, leading to discussions on how to effectively integrate both tools.

The conversation concluded with a consensus on the crucial need to understand potential bias, irrespective of the phase in which this task is incorporated. However, there was agreement that HR recruiters must be aware of bias risks before making any selection decisions. The importance of addressing bias at various stages of the recruitment process was emphasized, and the discussion underscored the significance of clarity in the integration of these functionalities within the Debiaser.

Effectiveness of the mitigation strategies:

The tool's mitigation strategies were generally well-received, with participants deeming them constructive. However, they also noted instances of the system providing inadequately explained and generic suggestions for bias mitigation. There was a common wish for the mitigation suggestions to be more pointed and helpful: many suggested focusing on turning perceived weaknesses into potential strengths and encouraging educational insights into why certain elements might be seen as bias-inducing. The participants also recommended that the system should incorporate more educational content about specific biases and their impact, understand the user's identity better for personalized suggestions, and use multiple language models for objectivity.

Fairness of the tool:

The second exercise brought forth various insights and discussions revolving around its potential hazards, trustworthiness, fairness, and scope for improvement. Participants acknowledged the usefulness of such a tool in eliminating bias from the recruitment process and the feedback in terms of fairness were relatively positive. However, they raised concerns about relying too much on its technology. They emphasized the importance of maintaining a degree of human critical thinking in the hiring process to prevent the potential risk of “placebic explanation” where the tool might generate a biased view of some candidates based on the information interpreted by it. Concerns were also raised around factual errors the tool might make due to user unfamiliarity with AI technologies, a wrong usage potentially leading to increased workload for HR.

Participants suggested the addition of an auditing process, perhaps by a third-party, as a way to certify the fairness of the tool. It was noted that a variety of thresholds, metrics, and evaluation grounds were necessary for fairness and the tool should be informed by country-specific regulations and laws. However, some argued that the artificial tool lacked an understanding of nuances and geography and that there existed an intrinsic distrust of AI compared to humans.

A few participants pointed out the inherent gender bias within the system and mentioned a rigid standardization in the outcome. Participants expressed the need for precision and favored traditional data manipulation methods over statistical ones for achieving optimal results. There was also a thorough discussion regarding the complex issue of whether to inquire about personal life, as it could potentially





serve as a proxy for social status and indirectly introduce bias. There was a debate about the concept of hiring based on motivations rather than experiences, as well as whether long-term tenure in a position could potentially introduce bias. The tool's ability to highlight positive biases was viewed as beneficial for broadening perspectives. There was a strong emphasis on the need for clear communication regarding a company's evaluation standards and the explicit disclosure of its diversity policies. However, the challenge of unwritten rules and the complexity of contextualizing standards within specific situations and cultures emerged as significant issues

Trustworthiness of the tool:

One of the most frequent hazards identified by participants was the possibility of over-reliance on technology, which might take away from human critical judgment in hiring processes. The tool could potentially increase bias in recruiters by pointing out biases they hadn't thought of, can be falsely used to further discrimination, and make factual mistakes that recruiters don't notice.

Users with no experience of AI might face difficulties and if used wrongly, the tool might increase work for HR, necessitating double-checks. It was stressed that user training and understanding were fundamental. Paradoxically, there were also GDPR compliance concerns raised about the processing of sensitive data required to train the system to operate fairly.

As already mentioned, suggestions for improving the trustworthiness of the tool included giving more in-depth, CV-specific feedback; individualizing bias explanations for each candidate; expanding on mitigation suggestions; and incorporating more educational functions based on previous studies, laws, and regulations.

General perceptions of AI:

Interestingly, the group made noteworthy remarks about the tool's behavior. The tool exhibited too much independence and underperformed compared to expectations, indicating a need for more stringent control. The participants noted instances of similarity and confirmation biases and queried the relevance of company experiences. The limitations of AI, particularly what was referred to as 'artificial dumbness', and the existence of historical data bias, were pointed out as weaknesses of such systems. Despite these shortcomings, AI's utility in documenting calculations was acknowledged positively, though simpler algorithms were suggested for transparency.

12.4 Second Block: Trustworthiness – Findings

12.4.1 Exercise Overview

After lunch it was held the second block of activities for the International Co-creation workshop. AI experts, HR professionals, and worker/CSO representatives collaboratively assessed risks and suggested solutions for integrating AI tools into recruitment systems. Every group was composed by experts in the field:

- **HR - KATHMANDU:**
 - 4 HR Specialists
 - 4 Partners
- **HR - HAVANA:**
 - 4 HR Specialists
 - 3 Partners
- **AI - MUMBAI:**
 - 4 AI Developers
 - 4 Partners
- **Worker/NGO - KYOTO:**
 - 4 Advocacy organization/NGO
 - 4 Partners
- **Worker/NGO - MARRAKECH:**
 - 4 Advocacy organization/NGO
 - 3 Partners

The method of commencing with hazards to derive solutions (future requirements) enabled participants to systematically examine critical issues identified in previous exercises and propose viable alternatives.





This approach showcased a commendable display of critical thinking. As anticipated, metrics emerged as the least explored theme due to its technical nature. Nevertheless, despite the technicality, the collaborative effort succeeded in gathering valuable insights from diverse sources.

The multi-perspective approach revealed a comprehensive set of risks and solutions, emphasizing the importance of transparency, fairness, and inclusivity in AI-driven recruitment systems. Coordination between HR and AI developers, along with ongoing assessment and adaptation, emerged as critical elements in addressing the identified challenges.

In paragraphs 9.4.2 and 9.4.3, key concerns and solutions are briefly outlined. Subsequent paragraphs provide a detailed list of risks, solutions, and actions for each category – AI experts, HR experts, and workers and NGO representatives. Each element is linked to the corresponding number of ALTAI's requirement relevant to that element, with some elements linked to more than one ALTAI requirement.

12.4.2 Key concerns over trustworthiness

Despite the enthusiasm generated by the simulation tools, several aspects raised concerns among participants. In the plenary session, the discussion over trustworthiness highlighted numerous risks, and some actions were already proposed, as observed in paragraphs 9.2.2 and 9.3.2. Building on these considerations, each group focused on its field of expertise to formalize the identified criticalities and transform them into ALTAI's sub-requirements.

Some participants raised concerns about the ALTAI tool's specifics. In particular, there were reservations about categorizing fairness as a subcategory of trustworthiness, given that a system could be fair yet lack user trust. On the other hand, other participants tended to interpret fairness as a synonym for trust, assuming that a truly fair system would eventually become trustworthy over time.

Concerns included the system's inability to handle a large number of candidates, hallucination risks from relying on incorrect information, and challenges in repeating experiments with consistent outcomes, as well as lack of understanding of the algorithm's reasoning, potential privacy issues hindering system robustness, and challenges in identifying and assessing soft skills; participants also identified de-responsibilization from human to machine, lack of human capability to supervise machine speed, and concerns about cybersecurity, privacy, and discrimination. Other risks involved potential discrimination based on sensitive attributes, the difficulty of HR users in understanding system explanations, the risk of cheating, hacking, information leakage, and the negative environmental impact of strong passive data collection. Accessibility issues, biases obtained by the tool, and a lack of public awareness were also pointed out.

12.4.3 Key solutions for trustworthiness

From an AI perspective, proposed solutions included the swift identification of unfit candidates, different strategies for different scenarios, and the incorporation of randomness in the selection process. Ensuring the system's awareness of sensitive data, making data sharing voluntary, and establishing a clear division between human and machine responsibilities were also highlighted. From an HR perspective, recommendations involved encouraging companies to permit employees to allocate time for public projects, advocating for an EU-level certification system, and the importance of education for users and stakeholders, which was highly emphasized. The requirements coming from workers representatives included human checks at critical stages, experimental bases for trustworthiness, security standards, and the exploration of an international standard. Measures such as opt-out options, third-party audits, and data masking techniques were suggested to address privacy concerns. Ensuring transparency, awareness, and inclusivity were key themes.





Actions across perspectives included the need for HR to double-check AI outcomes, conduct fairness checks, and ensure widespread education on AI systems. For AI developers, recommendations included involving HR testers in trials, configuring the ranking tool appropriately, and implementing a pilot and testing period.

12.4.4 ALTAI risks and solutions – AI Experts perspective

For clarity, the 7 ALTAI requirements titles are listed once again below:

- [1] Human Agency and Oversight
- [2] Technical Robustness and Safety
- [3] Privacy and Data Governance
- [4] Transparency
- [5] Diversity, Non-discrimination, and Fairness
- [6] Societal and Environmental Well-being
- [7] Accountability

Risks:

- Inability of the system to handle a large number of candidates [1]
- Risk of hallucination: even if the final system won't be based on LLM, relying on incorrect information is perilous in this context [2] [3]
- Oversight of important details [1] [2]
- Unreliable outcomes in repeating an experiment with the same input, resulting in different outputs [2][7]
- Uncertainty about candidates' willingness to share information [3]
- Awareness concerns regarding sensitive attributes for the system and individuals [3][7]
- Lack of sensitive data to train the model, posing a challenge in preventing discrimination [2]
- Difficulty for HR users in understanding explanations provided by the system [4]
- HR users facing challenges in linking machine decisions to practical input elements without supporting evidence [4]
- Potential discrimination based on sensitive protected attributes [5]
- Risk of discriminating against poor writers [6]
- Risk of cheating as individuals could pay others to write convincing motivation letters [6]
- Ambiguity in job offers, posing challenges in using them as input data [7]

Solutions:

- The model should swiftly identify reasons why a candidate is not a fit [1]
- Different strategies for different scenarios: filtering out excessive candidates while ranking those who make it to the shortlist [1]
- Ensuring the system is well-organized, clear, and concrete [1]
- Providing explanations of changes in filtering from different parameters and counterfactual scenarios [1]
- Incorporating randomness in the selection process [1]
- Ensuring the system is aware of sensitive data to avoid bias [2][3]
- Implementing blindness to sensitive traits for either humans or parts of the system [2][3]
- Conducting Counterfactual Testing [2] [3]
- Making data sharing voluntary, not mandatory [3]
- Citing text (CV and CL) for machine explanations and decisions [3]
- Establishing a clear division between human responsibility and accountability and the machine's role [2][7]

Steps and Actions:

- Actions for the HR:
 - HR persons should base their arguments on information from the model but retain the decision-making authority [4]





- The system should serve an inspirational purpose, leaving the final decision to HR [1][4][7]
- Ensuring knowledge of risks and awareness of the use of such tools [4]
- Disclosing the use of AI [7]
- Explaining the reasons and methods for using personal information without omissions [7]
- Clarifying the stage of the process where AI and humans are involved [7]
- Actions for the AI developers:
 - Additional functionality providing applicants with a description of why they are excluded from the process [7]
 - Configuring the system for social justice as an option [6]
 - Establishing benchmarks to demonstrate system performance [2][3][4][5][6]
 - Conducting Randomized Control Trials for the system [2][3][4][5][6]

12.4.5 ALTAI risks and solutions – HR Experts perspective

Risks:

- Lack of understanding among HR personnel regarding the reasoning behind the algorithm [1]
- Impediments to HR autonomy [1]
- Uncertainty about the allowed input from humans into the systems [1]
- Privacy concerns as a barrier to reinforcing system robustness [2][3]
- Variation in privacy issues based on the entity conducting recruitment, especially for public bodies handling sensitive data [2][3]
- Remove information from cv -> data masking -> but how about positive discrimination? [2]
- Legal challenges in identifying and assessing soft skills [3]
- Difficulty in establishing independent oversight due to challenges in identifying qualified parties, including potential use of social media information [1][2][3]
- Risks of system hacking and information leakage [2][3]
- Potential unlimited access from various internal and external sources [2][3]
- Challenges in comprehending system outputs [4]
- Balancing transparency to build trust without causing negative impacts [4]
- Tension between companies desiring clarity to avoid future explanations and the need for a structured application process for inclusion and objectivity, with caution against excessive objectivity [4][5]
- Concerns about previous decisions influencing the system in a biased manner [5]
- Highlighting biases and influencing those that may go unnoticed [5]
- The requirement of internet access for application leading to a digital gap [6]
- Potential surveillance concerns affecting worker perception and increasing stress [6]
- Negative environmental impact of strong passive data collection by systems [6]
- Lack of transparency in AI algorithms, making them less trustworthy [4][5][7]
- Ambiguity in responsibilities for development tasks [7]

Solutions:

- Encourage companies to permit employees to allocate time for public projects to enhance objectivity [1]
- Advocate for the establishment of a certification system at the EU level, funded by the government, to promote standardization [1]
- Implement data masking to remove personal information from CVs [3]
- Collaborate with data protection experts to develop practical processes for privacy preservation [3]
- Design a system that restricts access to personally identifiable information [3]
- Ensure repeatability and replicability of results to foster trust [4]





- Explore the existence of a certification for transparency [4]
- Enhance the structure of cover letters and applications [5]
- Promote education for users and stakeholders [5]

Steps and Actions:

- Actions for the HR:
 - Conduct a double-check of the AI outcomes [2][7]
 - Perform a fairness sanity check to assess the fairness of the process; HR should actively evaluate fairness [5]
 - Ensure education on AI systems is widespread across the community [6]
- Actions for the AI developers:
 - Include HR testers in the trial phase to ensure a comprehensive understanding of the tool [1][2]
 - Configure the ranking tool to create a combination of features fitting the roles [7]
 - Ensure proper testing and training for recruiters using the tool [7]
 - Implement a pilot and testing period for the tool [7]

12.4.6 ALTAI risks and solutions – Workers and CSO representatives Perspective perspective

Risks:

- De-responsibilization from human to machine, raising ethical and legal concerns [1]
- Insufficient human knowledge and capability to supervise the speed of machine operations [1]
- Non-determinism lacking trustworthiness without proper experiments [2]
- Cybersecurity risks, including concerns about data safety and storage [2][3]
- Safety considerations tied to control, questioning whether affected users need control [2]
- Excessive standardization potentially leading to errors and discriminatory outputs [2]
- Digital divide issues [3]
- Risk of data leakage [3]
- Uncertainty about GDPR-compliant data handling [3]
- Questions about data access, purposes of use, and availability [3]
- Importance of different forms of awareness for transparency to be effective [3][4]
- Consideration of traceability through logging certain actions for operationalized transparency [4]
- Balancing candidate visibility with concerns about how much visibility is too much [4]
- Lack of public awareness regarding how data is being utilized [4]
- Absence of feedback mechanisms for candidates [4]
- Accessibility issues leading to underrepresentation of minorities [5]
- AI finding proxies for discrimination, such as associating CV gaps with desertion without considering other factors like sickness or maternity leave [5]
- Risk of biases obtained by the tool confirming existing biases held by the user [5]
- Subjectivity in defining bias [5]
- Some discrimination grounds, like age and disability, may not always be considered discrimination, requiring clear definitions [5]
- Core expectations related to genuine and determining occupational requirements may pose challenges [5]
- Lack of regulations in place for AI systems [7]

Solutions:

- Human or a group of humans should check critical parts of the procedures [1]
- Implement milestone checks at critical stages of the process [1]
- Use experimental bases and statistical testing to enhance trustworthiness in non-deterministic scenarios [2][5]





- Establish security standards as a requirement [2]
- Explore the development of an international standard, potentially akin to ISO, to establish guidelines and benchmarks for assessing the trustworthiness of AI models [2]
- Provide an opt-out option, allowing individuals to choose not to have their application processed by AI [2]
- Certify security through third-party audits [2][3]
- Implement data masking techniques to protect sensitive information [3]
- Enable logging to allow visibility into different steps of the process [4]
- Increase awareness of how feature engineering is conducted [4]
- Consider adaptations for individuals with disabilities in accordance with local regulations [5]
- Utilize discrimination law as a basis for assessing the fairness of outcomes, aligning with international rights conventions [5]
- Include linguistic considerations in the development of the tool [5]
- Take into account the point of view of the candidates [5]
- Only include relevant information and clearly indicate the destination of irrelevant data [5]
- Conduct independent audits of AI systems [7]

Steps and Actions:

- Actions for HR:
 - Implement a communication mechanism, such as a letter, to highlight instances where AI may be making errors [1]
 - Simplify consent processes with GDPR-oriented, concise formats [2]
 - Provide a space for individuals to express disagreements [2]
 - Develop a checklist aligning with GDPR requirements [3]
 - Launch campaigns to inform the general public about how algorithms work with their data [3]
 - Implement Key Performance Indicators (KPIs) to regularly measure and compare results, and address user complaints a few months after the AI model's launch [5][7]
 - Establish and communicate policies and procedures [4][5]
 - Re-define qualities needed for a job, considering ethical and unbiased criteria [4][5]
 - Conduct third-party audits to ensure that candidates hired through the AI-driven recruitment process align with the requirements outlined in the job advertisement [5]
- Actions for AI developers:
 - Ensure human oversight in the development and deployment of AI systems [1]
 - Implement data masking techniques and ensure GDPR compliance to protect sensitive data [1][2][3]
 - Create a benchmark or standard for trustworthiness in AI methods [1]
 - Conduct audits using CEN/CENELEC standard verification audit processes [1]
 - Perform accuracy metrics auditing and random checks [3]
 - Implement measures to prevent data breaches [3]
 - Introduce red flags to prompt human intervention, potentially EU-based [1]
 - Develop a GDPR-assisting tool that highlights aspects requiring revision for compliance [4][5]
 - Implement Key Performance Indicators (KPIs) to regularly measure and compare results, and address user complaints a few months after the AI model's launch [5][7]
 - Establish metrics for assessing causality in AI methods [5]
 - Ensure robust data governance, advocate for modifications in international law, and promote open-source models [7]





13. Concluding remarks

This document presents the methodologies formulated by SVEN, in collaboration with BFH and NTNU, for the development of two rounds of national co-creation workshops and one international co-creation workshop, aimed at providing input for the early development phases of the Debiaser in W3. It also summarizes and provides analysis of the results from the two rounds of co-creation workshops conducted by seven partners in their respective countries between June and September 2023 as well as from the international co-creation workshop held in Venice in December 2023. As detailed in Chapter 4, the primary purpose of the first round of co-creation workshops was to identify, within the context of ad hoc prepared group work, sets of wordlists to be used by the technical partners of the consortium to feed T3.4.3 regarding 'bias' detection in training data. A total of 144 people from various stakeholder groups specified in the project's engagement strategy participated in the workshops, and partners collected 389 words/sentences that led to positive or negative bias as a result of the co-creative group activities.

Despite the inevitable differences among the workshops, a recurring bias was identified in all of them, particularly concerning the ethnic origin of candidates. Additionally, family situations were often viewed as potential sources of bias, especially in the case of female candidates. Potential bias arising from disabilities, sexual orientation, and the non-binary gender of candidates was also prominently noted in workshops that introduced these dimensions. However, out of the 389 words/sentences identified, only 38 were explicitly classified as causing gender bias, while 48 were related to race/ethnicity bias. A few were categorized as intersectional. Regarding the allocation among the different proposed categories, "career: work & education" had the highest number of words/ sentences, with 92, followed by "hobbies/leisure" (73) and "Family issues" (72). It's worth noting that the majority of words/sentences leading to negative bias were found in the "career: work & education" category (59 out of 92), while the opposite was observed in the "personal attitudes and other skills & knowledge" category, where the number of words/sentences leading to positive bias was almost double that of those leading to negative bias (40 out of 67).

The set of wordlists identified in the frame of the first co-creation workshop will support the work of the AI experts of the consortium in WP3. With an exploratory approach, a procedure has been outlined to more explicitly deepen the association/links between the individual words/sentences and the specific intersectional dimensions of inequality, also with the support from native speakers. This will allow for a more in-depth language and context specific understanding of the bias emerging from the wordlists. The different wordlists and sentence templates will be fed to existing methods to measure bias (e.g., WEAT), but for the word embeddings and language models in the local languages. On one side, this will give insights whether this real-world bias can be confirmed in the word embeddings and language models, and on the other side this enables the adaptation of the methods to measure bias to the specific challenges of the local languages and cultural aspects.

The focus of the second round of workshops shifted to examining the fairness of the recruitment processes, especially in the screening stage, and identifying desirable requirements and functionalities for a NLP tool and a CBR system, along with related risks, as it is explained in Chapter 6. In the implementation of these workshops, seven partners organized sessions engaging 131 participants to simulate a recruiting process and discuss principles of fairness in HR recruitment. The workshops, inclusive and well-attended, received positive feedback and identified 'Non-Discrimination' and 'Objectivity' as crucial fairness principles. Principles of objectivity and consistency were emphasized with discussions on eliminating personal bias through standardized criteria and structured interviews. The simulation of a hiring process for positions like Assistant Store Manager and Logistic Officer, revealed diverse criteria and challenges in achieving consensus. From a technical perspective, the observation of human reasoning and logic in the selection process was crucial for the development of a hiring system designed to emulate human decision-making (the CBR), and it laid the foundation for the algorithmic solutions to implement in the Debiaser.





The hands-on work aimed to identify AI tool requirements, revealing challenges in understanding NLP and CBR systems. Participants suggested streamlining initial screening, emphasizing unbiased filtering, and AI supporting decision-making rather than controlling it. Divergent opinions emerged regarding the role of AI, the definition of discrimination, and the inherent level of fairness within AI. Despite lacking broad agreement, the discussions enriched the understanding of fairness in recruitment, emphasizing the importance of a dynamic balance between diverse considerations.

From the second round of workshops it became evident that a more practical approach was necessary for stakeholders to give their relevant feedback on the needs and requirements for the Debiaser. Based on all the lessons learnt from the previous workshops, the international co-creation workshop was designed with specific methods and techniques that were derived from the previous ones. The aim of the third workshop was to further explore and deepen desirable requirements and functionalities for the NLP and CBR Debiaser systems. On the 7th of December 2023, in Venice (Italy), the international co-creation workshop was held, involving partners and external stakeholders: with 40 out of 45 registered participants actively engaged, the workshop comprised three blocks. The first block focused on simulating AI tools in recruitment, utilizing ChatGPT4 for interaction and fostering conversations around fairness and trustworthiness of an AI-driven solution. The Candidate Ranker methodology explored an hypothetical ranking phase, comparing human and machine decision-making. The Mitigation Tool delved into biases, simulating the word-embedding based Debiaser to mitigate biases in candidate evaluations. The second block involved theoretical discussions on trustworthiness, gathering technical requirements from AI, HR, and workers' perspectives, aligning with ALTAI requirements. The third block featured a collaborative brainstorming session addressing the learning needs of the various stakeholders involved, in order to contribute to the develop of a comprehensive training program in WP5.

The tech consultants of the Consortium were able to leverage on the findings and results of the workshop for the development of the proof-of-concept technology of the BIAS project, which is built in WP3. Having stakeholders from very different backgrounds, a user-centered development process covering the needs of different potential users was ensured. In particular, the collected feedback is relevant for the business case making use of the CBR component developed in WP3: the outputs from the international co-creation workshop, e.g., the relevance of values such as transparency, consistency, objectivity, job-relatedness, are directly influencing the design of this component. In particular, job-relatedness turned out to be of great importance and will be ensured in the phase of matching job announcements and CVs of the system. Furthermore, the findings from the workshops about what is perceived as sensitive (but not job-related) will influence the design choices of the CBR component.

The next step concerns the second co-creation phase which will notably have a different scope and goal: starting at M20 it will aim at shaping an exploitation path for BIAS in WP6. SVEN will develop a new methodology that will include periodic online discussion to take place on the Trustworthy AI Helix on the CrowdHelix [platform](#), in collaboration with FARPL, LEID, DIGI and CHX.





14. References

- [Ahn, J., Oh, A.. Mitigating Language-Dependent Ethnic Bias in BERT. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing \(pp. 533-549\). 2021](#)
- [Calistan A., Bryson J., Narayanan A.. Semantics derived automatically from language corpora contain human-like biases, Science 356, 2017](#)
- [Council of the European Union. Gender segregation in education, training and the labour market report by EIGE, 2017](#)
- [Council of the European Union. Gender segregation in education, training and the labour market Executive summary of the report by EIGE, 2017](#)
- [GE Academy project, D3.3 “Quality standards Booklet”, December 2021](#)
- [Eckstein D.. Styles of Conflict Management. Ottawa University, Sage Publications, 1998.](#)
- [Equinet. Domestic and care workers in Europe: an intersectional issue, 2022](#)
- [European Commission, Directorate-General for Research and Innovation, *She figures 2021 : gender in research and innovation : statistics and indicators*, Publications Office, 2021.](#)
- [Giancola O., De Vita L.. Between education and employment: women’s trajectories in STEM fields, Polis, 2017](#)
- [Greenwald A., McGhee D., Schwartz J., Measuring Individual Differences in Implicit Cognition: The Implicit Association Test, Journal of Personality and Social Psychology, 1998.](#)
- [Karla A. Henderson & Heather J. Gibson. An Integrative Review of Women, Gender, and Leisure: Increasing Complexities, Journal of Leisure Research, 45:2, 115-135, 2013.](#)
- [Kurpicz-Briki, M.. Cultural Differences in Bias? Origin and Gender Bias in Pre-Trained German and French Word Embeddings. 5th SwissText & 16th KONVENS Just Conference 2020, Zurich, 2020.](#)
- [Kurpicz-Briki M, Leoni T. A World Full of Stereotypes? Further Investigation on Origin and Gender Bias in Multi-Lingual Word Embeddings. Front Big Data. 2021](#)
- [Lee J-J., Jaatinen, M., Salmi, A., Mattelmäki, T., Smeds, R., & Holopainen, M.. Design choices framework for co-creation projects. International Journal of Design, 12\(2\), 15-31, 2018](#)
- [Leikas, Jaana & Koivisto, Raija & Gotcheva, Nadezhda. Ethical Framework for Designing Autonomous Intelligent Systems. Journal of Open Innovation: Technology, Market, and Complexity. 5. 18. 10.3390/joitmc5010018. 2019.](#)
- [Marchetti, S.. Care and Domestic Work. In: Migration and Domestic Work. IMISCOE Research Series. Springer, Cham, 2022.](#)
- [Mitchell S., Potash E., Barocas S., D’Amour A., Lum K., Algorithmic Fairness: Choices, Assumptions, and Definitions, Annual Review of Statistics and Its Application, 2021.](#)
- [Nielsen, Lene.. Personas in Co-creation and Co-design. Eleventh Danish HCI Research Symposium. 2011](#)
- [Nurski, L. and M. Hoffman \(2022\) ‘The impact of artificial intelligence on the nature and quality of jobs’, Working Paper 14/2022, Bruegel](#)
- [Robertson L., Abbas R., Alici G., Munoz A., Michael K.. Engineering-based design methodology for embedding ethics in autonomous robots, University of Wollongong, 2019](#)
- [Ruhl, E., Richter, C., Lembke, J., and Allert, H. Beyond methods: Co-creation from a practice-oriented perspective, A. \(eds.\), Design's Big Debates - DRS International Conference 2014, 16-19 June, Umeå, Sweden, 2014](#)
- [Saravanakumar K., The impossibility Theorem of Machine Fairness – A Casual Perspective, Columbia University, 2021](#)
- [Vlachopoulos, D., Makri, A. The effect of games and simulations on higher education: a systematic literature review. Int J Educ Technol High Educ 14, 22, 2017](#)
- [Wharton A., The sociology of gender: an introduction to theory and research, 2nd edition, Wiley-blackwell, 2011](#)
- [Waardenburg L., Huysman M.. From coexistence to co-creation: Blurring boundaries in the age of AI, Information and Organization, Volume 32, Issue 4, 2022.](#)
- [White, Hunter & Greaves. Facilitating Deliberation - A Practical Guide. MosaicLab, 2022.](#)





15. Annex 1 – Facilitation principles & conflict management tips

Participatory training principles rooted in feminist pedagogies

Facilitation principles and techniques integrated in the development of the present methodology that we suggest to apply to the BIAS co-creation workshops, are partly based on the principles of inclusive training developed in the frame of the [H2020 GE Academy project](#). The project, indeed, conceived the so called “PERFCKTSI” model, whose principles represent different angles and perspectives of inclusiveness, with the aim of contributing to promoting and practising social change by acknowledging and addressing exclusionary, power-based dynamics²².

The identified principles were applied to training and are selected and adapted to facilitation in co-creation settings as following:

1. **self-reflection and Reflexivity:** Both facilitators and participants constantly reflect the experience and the related learning process, acknowledging embedded power relations and reviewing their own practices and assumption.
2. **Contextualisation:** the process is context-specific and there is an effort to tailor it to the settings, situations, professional areas and needs of participants. This concerns all aspect, including contents, methods, materials and organisation.
3. **recognition of multiple “Knowledges” and relevance of “ownership” of knowledge:** Knowledge creation is regarded as a collective and inclusive process, and the diverse knowledges of participants are recognised, as well as how these are positioned differently. He co-design/co-creation process accommodates the sharing of the diverse knowledge owned by participants and facilitators.
4. **shared aim of social Transformation:** the co-creation process is not a stand-alone activity, but is intended as part of a broader social transformation strategy addressing unequal gender and diversity power relations.
5. **Standpoint awareness and critical perspectives:** co-creation contributes to make participants aware and respectful of the diversity of standpoints and identities which come into play in gender and diversity dynamics. At the same time, critical thinking is fostered, allowing to deconstruct these dynamics.
6. **Intersectionality:** co-creation supports participants in recognising and acknowledging the interplay of gender inequality and other forms of inequality and discrimination (including racism, xenophobia, classism, ageism, homophobia, transphobia and ableism) and to avoid homogenisation and binary conceptions of gender issues as well as interpretations of other forms of discrimination that do not take gender into account or are based on single-axis analysis that do to acknowledge the complex interconnections between gender, race, class, gender identity and sexual orientation, etc.

Facilitators’ role & responsibilities

“The facilitator needs to understand the group’s purpose, plan an appropriate process to achieve that purpose, lead the group through a range of activities, adjust the process to meet the needs of the group, intervene as needed to enable the group to resolve any problems and seek to ensure that the group achieves its purpose within the allocated timeframe” (White et al., 2022²³).

²² See [D3.3 “Quality standards Booklet”](#), GE Academy project, December 2021

²³ White, Hunter & Greaves. Facilitating Deliberation - A Practical Guide. MosaicLab: 2022





Leading facilitators have an intense, dynamic and adaptive role. They need to manage the whole workshop, paying attention to the group dynamics and making sure to keep the group focused on the topic.

The workshop's management includes introducing the workshop's agenda and all the different steps and activities, keeping track of the time and adapting the workshop's programme to possible time constraints in order not to exceed the allocated overall time.

They need to manage group dynamics and discussions, which will involve a range of communication skills, included summarizing the outcomes of the discussions.

They need to have a strong understanding of:

- BIAS's project objectives and workflow (for the introductory session);
- The topics for the panel discussion;
- The structure and purposes of the group work to be able to provide clear instructions.

They will need to "train" rapporteurs about their role, transferring the adequate knowledge, templates and documentation.

Facilitators are also responsible, supported by other staff members whereas needed, for:

- room set-up
- preparing, printing and distributing materials (provided links in this methodology)
- running microphones (during panel discussions)
- summarising outcomes of the panel discussions and the walking plenary.

Facilitation principles & tips

The following principles are inspired by the "Facilitating Deliberation - A Practical Guide" (White et al., 2022) integrated and adapted to the purpose of the co-creation group works in a different context than deliberative processes, taking the specific BIAS features into account.

- 1) **Comprehensive planning:** plan and conduct the process with consistency and ensure all participants understand and deliver their role;
- 2) **Independence and neutrality:** suppress personal views or emotional reactions, do not contribute with arguments to the discussion and avoid having decision-making authority;
- 3) **Clear purpose and task focus:** clear understanding and focus on the group work's tasks, giving enough time for their development;
- 4) **Respect for participants:** respect participants, supporting and encouraging them. Develop a climate of trust, behaving in a non-judgemental way;
- 5) **Respectful relationships among the lab's participants:** develop relationship-building activities and use moments of small-group work
- 6) **Participation:** encourage each participant to actively participate.

Conflict management

Due to the diverse categories of stakeholders involved in the workshop, disagreement might arise among participants, both in the frame of the panel discussion and within the group works. The event that disagreement generates conflict is quite unlikely due to the careful engagement strategy that is foreseen prioritizing stakeholders and individuals active in preventing and contrasting inequalities and/or with a pre-existing awareness on these matters. Still, in case this would arise, the tips for prevention and management highlighted below can be of use.





With conflict we mean both a “serious disagreement or argument, typically a protracted one”, and “struggle resulting from incompatible or opposing needs, drives or wishes”²⁴.

Conflicts can also originate from resistances of any stakeholders to specific topics. They can be manifested in two main ways:

- Active or explicit resistances include hostility, bad humour, devaluation and disparaging participants professional commitment, interrupting, etc. Other examples include the use of sexist/racist/homophobic language; and openly challenging the project methodology.
- Passive or implicit resistances are sometimes more difficult to identify. These include negative body language, foot dragging, inertia, chilly climate, making the procedures more difficult, giving less attention, uncomfortable social atmosphere, discomfort, inappropriate treatment, etc.²⁵

To prevent conflicts, a few tips can be adopted:

- Since the preparation phase, clarify with invited stakeholders which will be the specific topics for discussion, the activities and the expectations from participants.
- From the beginning of the workshop, encourage participants to consider different perspectives;
- Slightly adjust the agenda of the workshop if needed (for example, if the panel discussion takes more time than planned due to a conflict originated from different views on a topic).

About the resolution of conflicts instead, different “styles” can be adopted according to the specific situation (see table below, from Eckstein, 1998). It is recommended that, particularly if facilitators have no prior experience on this role, they dedicate some time to self-reflect on their own conflict management styles. We advise to follow the steps below for a short individual exercise:

- start from the “sentence(s)” column to reflect on which ones you feel better reflect your attitudes in tackling conflict;
- compare your choice with the described context when it is suggested to use that particular approach;
- try to envisage based on the expected participants to the BIAS workshop in your Lab to what extent conflict might arise and what strategies/approaches would suit: would you need to try and change your spontaneous attitude to tackle conflicts? How?

Table 33 Suggestions for conflict management in co-creation workshops

Sentence	Conflict management style	When to use it
I argue my case with participants to demonstrate the merits of the position I take.	Competing (highly goal-oriented, use aggressive behaviour to resolve conflicts, can be authoritative and uncooperative)	<ul style="list-style-type: none"> - When conflict resolution is urgent; when decision is vital in crisis - when conflicts involve personal difference that are difficult to change - When unpopular decisions need to be implemented

²⁴ <https://www.merriam-webster.com/>

²⁵ https://www.superaproject.eu/wp-content/uploads/2022/02/Resistances-to-Structural-Change-in-Research-and-Innovation_v02.pdf





<p>I seek to investigate issues with participants in order to find solutions that are mutually acceptable.</p>	<p>Collaborative (conflicts are seen as problems to be solved)</p>	<ul style="list-style-type: none"> - When maintaining relationships is important - When learning and trying to merge different perspectives - When time is not a concern
<p>I avoid discussing my differences with</p>	<p>Avoiding (better hide and ignore conflicts and resolve it, give up personal goals and display passive behaviour)</p>	<ul style="list-style-type: none"> - When confrontation will hurt a working relationship - When gathering information is more important than an immediate decision - When others can more effectively resolve the conflict
<p>I attempt to meet the expectation of participants</p>	<p>Accommodating (ignore own goals and resolve conflicts by giving into others)</p>	<ul style="list-style-type: none"> - When time is limited or when harmony and stability are valued - When suggestions/changes are not important to the accommodator - When maintaining the relationship outweighs other considerations
<p>I try to reach compromises through negotiation</p>	<p>Compromising (willing to sacrifice some goals while persuading others to give up part of theirs)</p>	<ul style="list-style-type: none"> - When important/complex issues leave no clear or simple solutions - When all conflicting people are equal in power and have strong interests in different solutions - When time is not a concern





16. Annex 2 – Scenarios proposed

1st scenario: Iron and steel industry looking for a warehouse worker

For permanent employment in a steel trading and processing company, we are selecting a warehouse worker with warehouse handling and loading/unloading duties.

The following are required:

- Container loading experience
- Qualification course for the use of forklifts and overhead cranes

Working hours 8-17.

Responsibility: Loading and unloading trucks. Moving and repositioning materials. Make sure that the production lines are always supplied with the materials necessary for their operation Goods handling. Use warehouse management software. Receive and manage shipments.

Employment contract: Full-time, Fixed-term, Permanent

Salary: €1,300.00 - €1,600.00 per month

Hours: From Monday to Friday

Types of additional pay: overtime, business bonus, thirteenth

Experience: Warehouse worker: 1 year (Required)

License or Certification: Excellent use of IT systems (Required)

2nd scenario: research institute looking for a Junior Group Leader

The Institute XXX is seeking outstanding, highly motivated candidates with an excellent scientific track record for a new Junior Group Leader position in the following areas of specialization:

- Bioengineering for Personalized Medicine (New diagnostic, modelling and prognostic systems to segment and identify the most appropriate treatment for each patient)
- Bioengineering for Advanced and Emergent Therapies (Bioengineering to develop therapies for human use based on genes (gene therapy)), cells (cell therapy) or tissues (tissue engineering) and including products of autologous, allogeneic or xenogeneic origin.

Applicants are expected to develop an ambitious project for their future group and to contribute to the center strategy based on excellent science, internationalization, translation and talent.

Candidates profile:

Apart from outstanding scientific output, the candidates must prove that they are active in the application of competitive proposals as principal investigators. Any mobility experience, e.g. a stay in another country/region, will be considered as a valuable contribution.

Desirable competencies and skills:

Leadership and people management; critical judgement in identifying and executing research activities; strategic vision for the future of the research field; income and funding generation; knowledge generation and transfer; collaboration; inclusion; excellent communication and networking.

What we offer:





The successful candidate will be appointed for an initial 4-year period. At the end of the fourth year, the Junior Group Leader will be evaluated by the International Scientific Committee. A positive evaluation will allow the candidate to extend their appointment for another 4 years.

Junior Group Leaders are offered a start-up package and provided with suitable laboratory space, access to the state-of-the-art core facilities and access to outstanding predoc students and postdoc researchers. Moreover, they are assigned a research project manager to support them with the management of their projects and interaction with friendly and researcher-oriented administrative staff. Emphasis is given to supporting their participation in competitive calls to start their own research line. Furthermore, the candidate adopts family-friendly policies to facilitate optimal work life balance for the successful candidate. Induction programme to facilitate incorporation and additional support is provided for foreigners to obtain Visa-working permit and to install in the city.

Furthermore, Junior Group Leaders have the opportunity to improve their career development through a wide range of professional training and coaching, and access to the international network of world-class research centres, universities, hospitals and industry. They also have the possibility to apply to calls for permanent group leaders.

3rd Scenario: Tech company looking for a Software engineer

We have an opening with a great client of ours and we are looking for a Software Engineer to work a Remote position. You must be a EU Citizen to apply. This is a yearlong contract position paying €35 to €75/hr. depending on your experience.

SOFTWARE ENGINEER

Designs, develops and has oversight for internal and external web pages and sites. Has responsibility for user interface, links, navigation flow, security and overall experience. May include creation of custom graphics and artistry. Maintains organization's communications strategies, message, branding and vision. May research new or related technologies.

This position will work closely with UI/UX designers and Front-end Developers and backend developers to create rich and engaging websites and applications for internal and external clients.

Some job responsibilities include

- Bring UI/UX designs to life using JavaScript and other code languages.
- Connect application front-end to data sources and web services using APIs
- Contribute ideas and perspective on team direction and technologies

Basic Qualifications

- Bachelor's degree or completion of certification program in web development.
- Experience with .net core (c#)
- Experience in a front-end technology and framework such as HTML, CSS, JavaScript, AngularJS, ReactJS, and JQuery
- Experience with virtualization/container software (Docker, or Amazon Web Services).
- Experience in writing SQL queries against a relational database
- Experience in REST and effective web service design
- Experience in a modern web application framework such as Ruby on Rails, Spring MVC, and Node.js

Desired Skills





- Bachelor's degree in computer science/technical discipline or completion of certification program in web development.
- Experience developing enterprise-level websites and applications.
- Experience with full-stack web development process.
- Experience developing unit tests and other quality assurance techniques.
- Some experience developing with WordPress.
- Excellent communications skills.
- Strong problem-solving techniques.
- Demonstrated dedication to creating positive client experiences.
- Passion and curiosity for new technologies.
- GitHub's profile with submissions to open source projects

4th scenario: a private school looking for an educator

We are currently seeking to appoint experienced, creative and dynamic Early Years Educators who are fully qualified to work with children from 3 to 6 years of age; educators who will be able to adapt quickly to a new and exciting learning environment, educators with strong empathy skills who are willing to use relational strategies in a stimulating learning environment with students from different cultural and linguistic backgrounds.

A strong commitment to purposeful learning and teaching, collaborative planning and open communication is essential. We pride ourselves on the commitment offered by our friendly, supportive and professional staff. Applicants must be willing to be active, flexible participants in a hard-working team of teachers. Proven experience within an international, bilingual or multilingual context would also be useful.

Due to the expansion of the school, we are currently seeking to appoint experienced, creative and dynamic teachers of the highest calibre to join us from September 2023. Potential teachers must possess the following experience, skills and knowledge:

- Bachelor's Degree in Early Years specifically
- Established experience in teaching in the Early Years - No NQTs
- Be an English first language user with excellent oral and written communication skills
- Demonstrate an ability in engaging with students and teaching staff alike, maintaining a high standard of work at all times
- Excellent interpersonal and organisational skills
- Demonstrate enthusiasm, commitment and professionalism at all times
- Have a positive and flexible approach to school life and a well-developed ability to work in teams
- The ability to create a happy, challenging and effective learning environment
- Hold "Qualified Teacher Status" or an equivalent qualification
- Have a minimum of two years' teaching experience
- Knowledge of the EYFS is an asset but not absolutely essential

Initial interviews will take place by Zoom. Please note that the interview process may begin earlier than the closing date for applications.

Flights, baggage allowance and support for an apartment search are provided by the school.

Candidates should upload their letter of presentation, Curriculum Vitae and a recent photograph when submitting their application.

Please send your email with the relevant position as your subject line.





Our school is committed to safeguarding and promoting the welfare of all the students and we expect all applicants to share this commitment. We ensure that safe recruitment practices are followed and hold ourselves accountable to the highest standards. All appointments will be subject to an interview, criminal record checks and two successful references.

We thank all applicants for their interest in this role, however, please be advised that only applicants shortlisted for an interview will be notified.





17. Annex 3 – Personas for the first workshop

Jamal



Born in Ghana on the 13.10.1985

Man, cisgender, Married with 3 children

Previous work experiences:

.....
.....

Education:

.....
.....

Hobbies/sports and personal attitudes:
yoga teacher, team building attitude,

.....
.....

Camila



Born in Argentina on the 20.05.1995

Woman, cisgender, Single

Previous work experiences:

.....
.....
.....

Education:

.....
.....
.....

Hobbies/sports and personal attitudes:
cycling & running, leadership skills ,

.....
.....



Lana



Born in Vietnam on the 15.02.2000

Woman, trans, Single

Previous work experiences:

.....
.....
.....

Education:

.....
.....
.....

Hobbies/sports and personal attitudes:
tennis, netflix series passionate

.....
.....

Ahmed



Born in Iran on the 3.11.1993

Man, cisgender, married with 2 children

Previous work experiences:

.....
.....
.....

Education:

.....
.....
.....

Hobbies/sports and personal attitudes:
loves pets, horse riding, good communication skills

.....
.....



Mona



Born in Egypt on the 1.09.2003

Non-binary, engaged no children

Previous work experiences:

.....
.....
.....

Education:

.....
.....
.....

Hobbies/sports and personal attitudes:
gaming, problem solving attitude,

.....
.....

Selina



Born in Italy on the 3.08.2000

Woman, cisgender, Roma, single, 1 child

Previous work experiences:

.....
.....
.....

Education:

.....
.....
.....

Hobbies/sports and personal attitudes:
drawing, playing guitar in a group, management skills,

.....
.....



Mainstream profile Born in on the

picture

.....

Previous work experiences:

.....
.....

Education:

.....
.....

Hobbies/sports and personal attitudes:

.....
.....





18. Annex 4 – template of cover letter

Cover letter





19. Annex 5 – Template for the walking plenary

BIAS

Mitigating biases of AI in the labour market

Iron and steel industry looking for a warehouse worker			
Job offer	<table border="1" style="width: 100%;"> <tr> <td style="width: 50%; padding: 5px;">Job offer main features, expected skills and competences:</td> <td style="width: 50%; padding: 5px;">Potential bias deriving from the job offer's formulation</td> </tr> </table>	Job offer main features, expected skills and competences:	Potential bias deriving from the job offer's formulation
Job offer main features, expected skills and competences:	Potential bias deriving from the job offer's formulation		
Fictional character	<table border="1" style="width: 100%;"> <tr> <td style="width: 50%; padding: 5px;">Fictional character main features (gender, race, age, marital status)</td> <td style="width: 50%; padding: 5px;">Fictional character CV (previous work experiences, education, hobbies/personal attitudes)</td> </tr> </table>	Fictional character main features (gender, race, age, marital status)	Fictional character CV (previous work experiences, education, hobbies/personal attitudes)
Fictional character main features (gender, race, age, marital status)	Fictional character CV (previous work experiences, education, hobbies/personal attitudes)		
Cover letter	<table border="1" style="width: 100%;"> <tr> <td style="width: 50%; padding: 5px;">Words/sentences of the cover letter (nouns and adjectives) at risk of bias</td> <td style="width: 50%; padding: 5px;">Words/sentences rephrased</td> </tr> </table>	Words/sentences of the cover letter (nouns and adjectives) at risk of bias	Words/sentences rephrased
Words/sentences of the cover letter (nouns and adjectives) at risk of bias	Words/sentences rephrased		
Other relevant outcome of the work			





20. Annex 6 – Reporting template for note takers – first workshop

Date, Place

Group/scenario number	
Group composition	
<p>First activity - discussion on the job offer (the note taker reports on each proposed point of discussion and on potential bias deriving from the job offer’s formulation, highlighting critical words/sentences) - 15 minutes</p>	<ul style="list-style-type: none"> • which are the prerequisites/expected skills and competences • which is the ideal profile according to the offer text • which elements she/he would expect to find in a successful cover letter • potential bias deriving from the job offer’s formulation
<p>Second activity - elaboration of the cover letter (the note taker reports on the main topics of the discussion of the group, on which points the group focused at most, if different points of view emerged) - 30 minutes</p>	
<p>Third activity - discussion on the cover letter (the note taker takes notes for each question proposed paying attention to indicate if a identified bias is positive or negative - use different colors - green for negative and blue for positive) - 20 minutes</p>	<ul style="list-style-type: none"> • Are there any risks of bias coming from the cover letter? • which kind of bias (e.g. gender, race, age, disability, etc.)? • Which are the words/sentences that could lead to bias? which are the associations to those words/sentences that make them lead to bias?





	<ul style="list-style-type: none"> • How can these risks affect the decision of a recruiter? • Are the biases related to the work or family/private life sphere or any other category?
<p>Fourth activity - cover letter rephrasing (The note taker reports in the template provided by facilitators notes on main arguments/topics of the discussion of the group, on which points the group focused at most, if any different points of view emerged) - 20 minutes</p>	
<p>Please indicate if and which different positions have emerged during the discussions above according to the different stakeholder categories. In particular, indicate if different stakeholders positioned/focused on different kinds of bias.</p>	





21. Annex 7 – First workshop report

Report first co-creation workshop

Date, place

Partner:

Participants

Categories	Number
HR officers	
Representatives of HR networks	
AI specialists	
Workers	
Workers' representatives	
Representatives of NGOs, networks, organisations fighting against discriminations	
Other	

1st part - Results of the open discussion

1. Please report how you structured the open discussion and if you followed the 5 questions proposed in the methodology and if not which statements were brought to the attention and which questions were posed to the panel
2. Were participants engaged in the discussion? Did any particular dynamic/tension among the different categories of participants emerge?
3. Can you summarize the main point of discussion related to the 5 questions proposed? (pay particular attention to results of question and different definitions of fairness)

Question 1: <i>based on your knowledge and experience, do we have good reasons to be optimistic or should we rather be concerned with the situation in our country?</i>	
Question 2: <i>what is your opinion and/or experience as far as the use of AI systems in recruitment and Human resources management in general? Are the PROs</i>	





<i>overly tech-innovation enthusiasts? What points of attention would you advise to balance the CONs?</i>	
Question 3: <i>what is your view on the role that AI based technology can play to favour or to hamper EDI in hiring processes in particular?</i>	
Question 4: <i>how would you define it and to what extent such definition is context dependent in your view?</i>	
Question 5: <i>what is your view on this? How participation of workers and social partners but also civil society organizations representing minorities can contribute to influence and oversee the use of AI in recruitment and make it fairer?</i>	

2nd part - Results of the group works

N. of groups created	
Groups' composition (please specify the number of participants per each category e.g. workers, HR officers, AI specialists, etc.)	
Scenario & fictitious characters (please indicate which fictitious characters and which scenario you have used. <u>In case you elaborated a different scenario and/or focused on a different ground of discrimination than race/ethnicity, please explain why you chose such specific scenario and discrimination dimension. Also, as regards the gender dimension please specify if you have opt for binary or non-binary profiles and why</u>)	

1. Please provide with a summary about how the group work developed, explaining if any particular dynamics or issues emerged.
2. Was the collaboration among the different stakeholders smooth and contribution balanced?
3. Which one of proposed activities resulted more easy/smooth to implement and which one resulted to be more difficult? Why?



4. Can you please synthesize the contents and dynamics of the discussion in the groups for each activity?

Activity	Contents and dynamics of the discussion
First activity - discussion on the job offer	
Second activity - elaboration of the cover letter	
Third activity - discussion on the cover letter	
Fourth activity - cover letter rephrasing	

5. Can you please summarise the categories of bias as well as words/attributes/sentences considered to be potentially biased by participants that emerged in the different activities proposed? Please report both positive (in blue) and negative (in green) bias?

Activity	Categories of bias (career, family issues, work ethics, Personal attitudes and other skills & knowledges, Hobbies/leisure, other categories)	Words/sentences
First activity - discussion on the job offer		
Second activity - elaboration of the cover letter		
Third activity - discussion on the cover letter		
Fourth activity - cover letter rephrasing		

6. Did the different categories of participants in the workshops (HR officers, workers, AI specialists) happen to focus on specific bias and related words? If so, which ones?
7. Were specific actors found to be conveyors of specific bias?
8. Please report on the final plenary walking session. Did anything particular emerge?
9. Which aspects, topics of discussion, bias, emerging from the workshop could be identified as country specific or peculiar to the cultural/socio-economic context ? Can you give some examples?



22. Annex 8 - Material for the second round of workshops – first group work

Candidates Profiles for first group:

Candidate 1 (he)



Born in Germany on the 1.09.1995

Single

Previous work experiences:
2 years experience in the retail sector (assistant store manager and logistics officer)
3 years experience in production quality control

Education:
Bachelor Degree in Economics
Technical Institute diploma

Hobbies/sports and personal attitudes:
Yoga and meditation teacher, leadership skills, empathy and attention towards the customer, foreign languages knowledge

Candidate 2 (she)



Born in India on the 03.05.1990

Married, 2 children

Previous work experiences:
5 years experience in food manufacturing
2 years experience in retail sector
3 years experience in as store manager in supermarket

Education:
High School diploma in Industry-technical institute

Hobbies/sports and personal attitudes:
Gaming, problem solving and team working attitude, Microsoft Office knowledge

Candidate 3 (they)



Born in Italy on the 18.01.2000

Single

Previous work experiences:
3 years experience in retail as store manager
1 year experience as intern junior translator

Education:
MA student in Industrial Engineering
Bachelor Degree in Engineering
High school: linguistic/foreign languages

Hobbies/sports and personal attitudes:
Basketball, food influencer on Instagram, team working attitude, computer skills

1st Job offer

Assistant store manager at XYZ

XYZ, Padova, Veneto, Italy

Introduction

As assistant store manager you will collaborate with the Store manager for the store commercial and economic management. If you have communicative skills and proactivity this is the right job offer for you!

The Job Position

Thanks to an initial period of training, you will have deep knowledge about all the activities that fall into this role. In compliance with company policy and in accordance with the store manager, you will be a reference point for all the store employees.

The main tasks for this Job Offers are:





- Personnel management defining work shifts and organizing training activities for shop assistants
- Checking both prices and all displayed range of product
- Care of the shop appearance (cleanliness and order)
- Ensuring compliance with the law about personnel security and food security (HACCP)
- Client assistance
- Store Manager substitution when they is absent

The desired profile

The requirements to cover the job offer as assistant store manager are:

- High School Diploma or University Degree
- Precedent experience in retail or in large scale retail trade
- Time flexibility and dynamicity
- Team working attitude to achieve common goals
- Strong client orientation

What we offer:

A dynamic job organized on weekly work shifts distributed in about 5 days, but possibility to work also during the weekend according to need. You will be able to give a fundamental contribution to the Store Manager in the commercial economic management and in the store organization. We offer a full time or part time permanent employment contract with an interesting remuneration package starting from € 26.650 (full time), concrete career prospects possibility and economic advancement during the following years. We also offer interesting benefits provided by the company welfare system.

Company profile

XYZ is a leading company in the large scale retail trade and presents a dynamic context which has been certifying Top Employer for several years.

XYZ has adopted a gender & diversity plan to ensure an inclusive workplace and it has put in place concrete actions to guarantee more women and minority representation in the team.

XYZ has adopted a protocol to ensure an optimal life-work balance for its employees, which includes the possibility of horizontal and vertical part time collaboration, as well as time flexibility.

The company offers to employees a complete and structured onboarding programme to learn about the company, the role and the challenges and that alternates e-learning modality training and training on the job, as well as continuous training during all the career path with the training manager constant support.





Candidates Profiles for second group:

Candidate 1 (he)

Born in Germany on the 1.09.2000

Single

Previous work experiences:

1 year experience in logistics for a big retail company
1 year experience as cashier in XYZ supermarket
1 year experience as cashier in supermarket

Education:

Technical Institute diploma

Hobbies/sports and personal attitudes:

Yoga, empathy and attention towards the customer, foreign languages knowledge, live closed to the work place

Candidate 2 (she)

Born in India on the 03.05.1990

Married, 3 children

Previous work experiences:

5 years experience in logistics
5 years experience as sales assistance

Education:

High School diploma in Industry-technical institute

Hobbies/sports and personal attitudes:

Gaming, problem solving and team working attitude, very flexible working hours, live close to the work place

Candidate 3 (they)

Born in Italy on the 18.01.1986

Married

Previous work experiences:

10 years experience in retail sector as logistic operator

Education:

Bachelor Degree in Engineering
High school: linguistic/foreign languages

Hobbies/sports and personal attitudes:

Cooking, spending time with my children, team work, very precise

*2nd Job offer***Logistics Officer at XYZ****Job offer overview**

The reference area is the Logistic Department. The Logistic Officer is expected to work within our Distribution Centre.

Responsibility of the role

The Logistics Officer operates within the warehouse and must ensure the proper execution of planned activities. The main task is to monitor and manage the regular inspection of incoming and outgoing goods.

Activities

The resource will be responsible for:

- Ensuring compliance with company quality standards by conducting meticulous checks regarding expiration dates, package integrity, barcode legibility, size, pallet and packaging integrity.
- Verify the correspondence between received items/quantities and placed orders to meet the needs of retail outlets.





- Once the checks are completed, enter the required information into the computerized management system.

Onboarding and Training

After a period of working alongside warehouse personnel and on-the-job training, the resource will be fully capable of carrying out the assigned activities.

Candidate Characteristics

Requirements:

- High school diploma (five-year program)
- Residence or domicile in areas near our Distribution Center
- Availability to work in shifts, including night shifts.

Personal Characteristics:

- Team working
- Precision and reliability
- Flexibility

Company Profile

XYZ is a leading company in the retail industry.

XYZ has implemented a recruitment protocol that prioritizes candidates with previous work experience in XYZ, even in different countries.

XYZ has not yet adopted a gender & diversity plan. XYZ has also implemented a protocol to ensure the well-being of its employees and has recently approved a “retention policy” with a set of guidelines to ensure that employees are satisfied with their work, in order to motivate them to stay in the long term.





23. Annex 9 – Template report first group work – second workshop

Discussion in two groups_Report group 1

Group composition:

Stakeholders' kind	Number
HR officers	
Workers & workers' representatives	
AI specialists	
Philosophers	
Legal experts	
Representatives of CSOs	

Job offer and company's profile:

Kind of job offer	Company's profile used - features

Candidates' profiles:

Candidate 1	
Candidate 2	
Candidate 3	

1st part - discussion in two groups

1st activity

Please report the result of the Mentimeter poll (you can also attach a screenshot of the results with translation in English)

- Which principle received more votes when asking which one is more important?
- Which principle received more votes when asking which one is less important?





- Please briefly report on what emerged from the following discussion on the poll.

2nd activity

Please report on the discussions regarding the application of the two principles of objectivity and consistency

Principle	Procedures, measures, tools, data/info used to implement them
Objectivity	
Consistency	

other notes:

3rd activity

Report on the discussion and the specific questions posed by the facilitators:

- Reflect on the information /variables provided both regarding the candidates and the company profile: which ones do you consider important to take into account in a first screening of candidatures received? Can you agree on an order of importance?
- Reflect on elimination criteria: among the 3 candidates is there one candidate that you would eliminate for sure? If yes which one? Why and how did you get to this decision? Which is the reasoning behind?
- Reflect on the selection criteria: if you have to chose a candidate to interview among the 3, which one would you interview? Why and how did you get to this decision? Which is the reasoning behind?
- (In case it did not emerge during the conversation, explore) how much important were the company's rules in orientating the decision on the candidates to eliminate and to invite to an interview?

Also elaborate on the following questions:

- Was there consensus among HR officers?
- Which questions triggered different reflections/considerations?
- Which were the more diverging reasonings/reflections?
- Was a compromise found or couldn't participants reach a final common decision on the different questions?

Report on the plenary:

- Were major differences in the approaches and perspectives from the 2 groups reported and/or discussed in the plenary?
- Which were the main feedbacks from the other participants with observer roles?

Kind of stakeholder	
Legal experts	
AI specialist	





Philosophers	





24. Annex 10 – Table for the fourth activity – second workshop

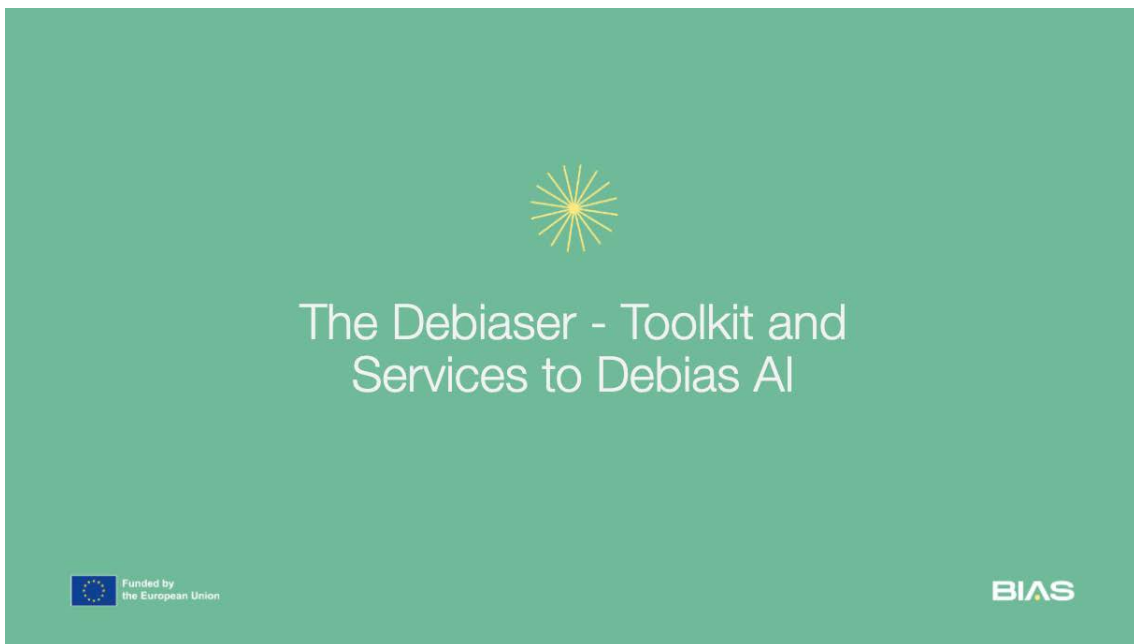
Group work - Fourth activity

Requirements	Conditions for being considered fair/trustworthy	How to evaluate fairness and trustworthiness
e.g. CVs screening highlighting any sensitive information/expression in the CVs that are at risk of causing bias	e.g. the system should take into account minorities e.g. the selection of the group of candidates to be interviewed is strictly related to the professional experience and not to sensitive attributes	e.g. the result of the screening made by the AI system should present a diverse group of candidates with the 80% rule: for every 5 white men invited, invite 4 from minority group e.g. the results of the screening should include an explanation for each candidate selected for interview by the system e.g. the system should present a score of the accuracy of the explanations of the information/expressions at risk of bias





25. Annex 11 - Slides introducing the Debiaser and the CBR model





The Debiaser - Toolkit and Services to Debias AI

- Debiasing and explaining text-based recruitment applications, incorporating principles of Fair and Trustworthy AI
- Target Group - HR: Companies using digital tools for recruitment
- Additional key stakeholders:
 - Employees/Candidates
 - Regulators/Auditors

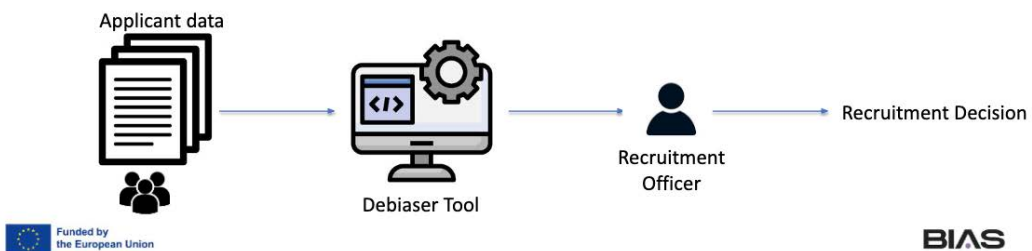


BIAS



The Debiaser Tool

- The tool processes applicant data (focus on text) with aid of debiased AI models.
- A recruitment officer receives the output of the tool and makes a decision based on that information.
- MAIN QUESTION: In order to promote fair, trustworthy, unbiased AI-aided recruiting...
 - What kind of data should the tool process?
 - What should be considered in the tool's design?
 - What should the tool present to the user?



BIAS





The Debiaser - Sample functionality ideas



BIAS

Biased text-based AI in Recruitment?



Whereas data in texts such as motivation letters in applications are anonymized to avoid bias, this is often not enough.

Some patterns not explicitly visible to the humans might be picked up by the AI and lead to unfairness in decision making upon such documents.

This is in particular delicate for the matter of automated recruitment tools.



BIAS

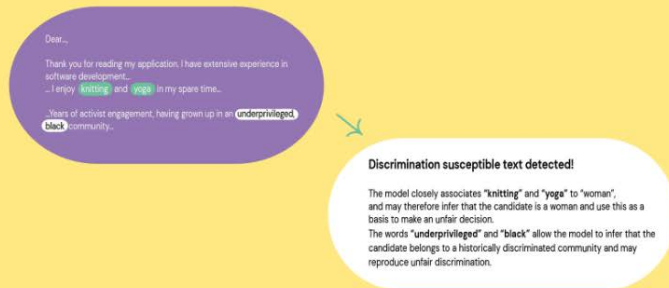




The Debiaser shows you which parts of the application might be subject to discrimination and how to deal with these parts in the form of bias mitigation.

Different scenarios can be tested to verify the different classification outcomes based on the proposed mitigation measures.

BIAS DETECTION – INPUT SIDE



(illustrations are an early vision and the final product might be subject to changes)



It furthermore provides human-understandable explanations of the parts that impacted the automated decisions you rely on, building trust in them.

Along with the technical tools to achieve this, know-how sourced from European research excellence will be delivered in form of consulting.

DEBIASED RECRUITMENT TOOL
(USING DE-BIASED MODEL)



(illustrations are an early vision and the final product might be subject to changes)





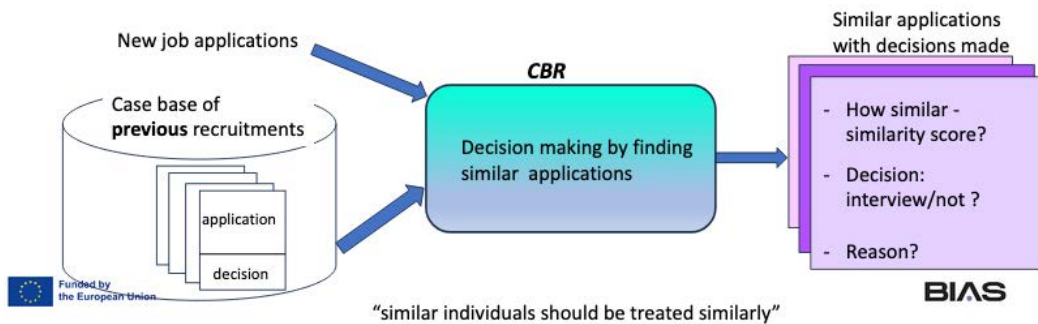
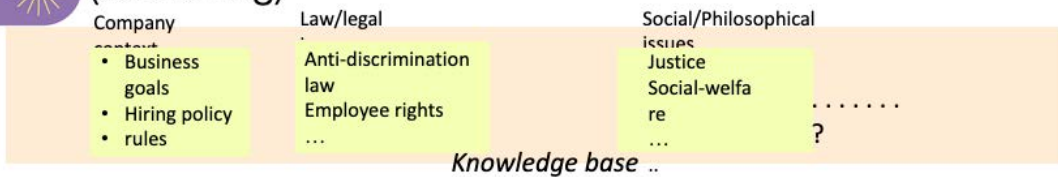
Other functionalities under exploration



- Provide explanation of decision based on EU-non-discrimination provisions, job description, company policies and candidate
- Flagging and explaining potential triggers of unconscious bias in recruitment officer.
- Adjusting for different preferences of user incorporating DEI, suitability, uniqueness etc.

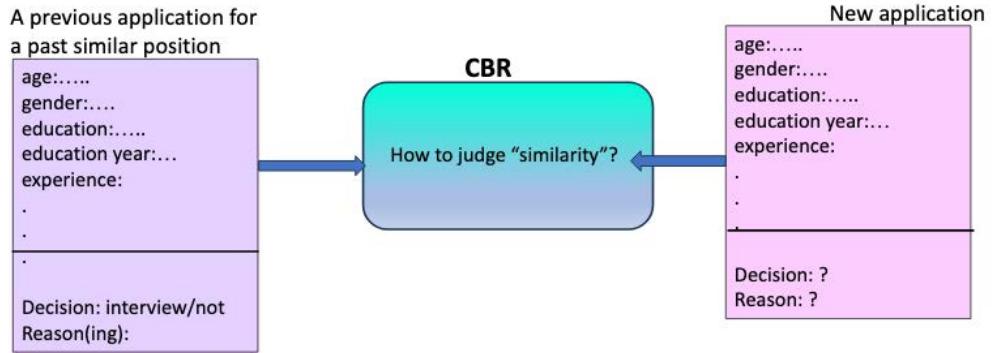


CBR as decision support for fair recruitment (screening)





Fair decision support through re-using past "similarity" experiences





26. Annex 12 – Second group work report – second workshop

Date, Place

Group n.		
Group composition		
First activity – brainstorming on the screening phase – 10 minutes	<p>How could an innovative technology based on Natural Language Processing (NLP) and Case Based Reasoning (CBR) support the three steps of the screening phase? (try to differentiate and/or link each suggestion to one or more of the 3 steps)</p> <p>Which needs would the technology address?</p> <p>Diverging positions/ideas among participants</p>	
Second activity – identifying requirements - 20 minutes	<p>What should the tool do? (requirements) (Please try to differentiate and/or link each requirement to one or more of the 3 steps)</p> <p>Requirement 1:</p> <p>Requirement 2:</p> <p>Requirement 3:</p> <p>.....</p>	<p>How should the tool be to do it? (Please try to differentiate and/or link each suggestion/feature to one or more of the 3 steps)</p>





	Diverging positions/ideas among participants		
Third activity - identifying inputs and outputs of the tool - 15 minutes	Requirements (Please try to differentiate and/or link each requirement to one or more of the 3 steps)	Inputs	Outputs
	Requirement n. 1		
	Requirement n. 2		
	Requirement n. 3		
	Requirement n. 4		
	Diverging positions/ideas among participants		
Fourth activity - identifying conditions for fairness/trustworthiness and evaluation 15 minutes	Conditions for being considered fair/trustworthy	How to evaluate fairness and trustworthiness	





	Diverging positions/ideas among participants	
Walking plenary - 30 minutes	Reports any interesting outcomes coming from the plenary sessions that did not emerge during the group work itself	





27. Annex 13 – Overall report – second workshop

Overall report 2nd co-creation workshop

Date, place

Partner:

Participants

Stakeholders' kind	Number
HR officers	
Workers & workers' representatives	
AI specialists	
Philosophers	
Legal experts	
Representatives of CSOs	

1st phase - Discussion in two groups

Group 1 composition:

Stakeholders' kind	Number
HR officers	
Workers & workers' representatives	
AI specialists	
Philosophers	
Legal experts	
Representatives of CSOs	

Group 2 composition:

Stakeholders' kind	Number





HR officers	
Workers & workers' representatives	
AI specialists	
Philosophers	
Legal experts	
Representatives of CSOs	

Job offers and company's profiles:

Groups	Kind of job offer	Company's profile used - features
Group 1		
Group 2		

Candidates' profiles:

Groups	Candidates	Description
Group 1	Candidate 1	
	Candidate 2	
	Candidate 3	
Group 2	Candidate 1	
	Candidate 2	
	Candidate 3	

1st activity

Please report the result of the Mentimeter poll of the two groups (you can also attach a screenshot of the results with translation in English)

- Which principle received more votes when asking which one is more important?





- Which principle received more votes when asking which one is less important?
- Please briefly report on what emerged from the following discussion on the poll

2nd activity

Please report on the discussion of the groups regarding the application of the two principles of objectivity and consistency

Principle	Procedures, measures, tools, data/info used to implement them
Objectivity	
Consistency	

other notes:

3rd activity

Report on the discussion and the specific questions posed by the facilitators. Make a summary of the results of the discussion in the two groups:

- Reflect on the information /variables provided both regarding the candidates and the company profile: which ones do you consider important to take into account in a first screening of candidatures received? Can you agree on an order of importance?
- Reflect on elimination criteria: among the 3 candidates is there one candidate that you would eliminate for sure? If yes which one? Why and how did you get to this decision? Which is the reasoning behind?
- Reflect on the selection criteria: if you have to chose a candidate to interview among the 3, which one would you interview? Why and how did you get to this decision? Which is the reasoning behind?
- (In case it did not emerge during the conversation, explore) how much important were the company’s rules in orientating the decision on the candidates to eliminate and to invite to an interview?

Also elaborate on the following questions:

- Was there consensus among HR officers?
- Which questions triggered different reflections/considerations?
- Which were the more diverging reasonings/reflections within the groups and between the two groups?
- Was a compromise found or couldn't participants reach a final common decision on the different questions?

Report on the plenary:

- which were the main feedbacks from the other participants

Type of stakeholder	
Legal experts	
AI specialist	



Philosophers	

2nd phase - group work in 4 groups

N. of groups created	
Groups' composition (please specify the number of participants in each group per category e.g. workers, HR officers, AI specialists, etc.)	

1. Please provide with a summary about how the group work developed, explaining if any particular dynamics or issues emerged.
2. Was the collaboration among the different stakeholders smooth and contribution balanced?
3. Which one of proposed activities resulted more easy/smooth to implement and which one resulted to be more difficult? Why?
4. Can you please synthesize the results of the discussion in the groups for each activity? Please report the overall results of the groups for each point

First activity - brainstorming on the screening phase - 10 minutes	<ul style="list-style-type: none"> • How could an innovative technology based on Natural Language Processing (NLP) and Case Based Reasoning (CBR) support the three steps of the screening phase? <ul style="list-style-type: none"> • step 1 • step 2 • step 3 • Which needs would the technology address? • Diverging positions/ideas among participants 	
Second activity - identifying requirements - 20 minutes	<ul style="list-style-type: none"> • What should the tool do? (requirements) 	<ul style="list-style-type: none"> • How should the tool be to do it?
	step 1 1) 2)	step 1
	step 2	step 2
	step 3	step 3



	Diverging positions/ideas among participants		
Third activity - identifying inputs and outputs of the tool	Requirements	Inputs	Outputs
	Step 1		
	Step 2		
	Step 3		
	Diverging positions/ideas among participants		
Fourth activity - identifying conditions for fairness/trustworthiness and evaluation	Requirements	Conditions for being considered fair/trustworthy	How to evaluate fairness and trustworthiness
	Step 1		
	Step 2		





	Step 3		
	Diverging positions/ideas among participants		
Walking plenary	Please report on the final plenary walking session. Did anything particular emerge?		





28. Annex 14 – Simulation tools material

Job offer:

Job Title: Assistant Store Manager/Salesperson at XYZ

Location: Olten, Switzerland

Job Description:

As an Assistant Store Manager at XYZ, located in Olten, Switzerland, you will play a crucial role in the commercial and economic management of a flagship franchise within a supermarket chain. Collaborating closely with the store manager, you will be responsible for overseeing various aspects of store operations, ensuring a positive customer experience, and maintaining legal compliance in personnel and food security. This dynamic position involves personnel management, pricing and product range oversight, visual merchandising, and stepping in as the acting store manager when required.

Minimum Qualifications:

- Completed apprenticeship in the retail trade or equivalent experience.
- Experience in retail or large-scale retail trade.
- Time flexibility to accommodate weekly work shifts, including weekends.
- Excellent teamwork and social skills.
- Strong customer orientation.

Nice-to-Have Qualifications:

- Previous experience in a supervisory or assistant management role.
- Familiarity with large-scale retail trade operations.
- Additional training or certification in retail management.

About the Company:

XYZ is a leading company in the large-scale retail trade, situated in Olten, Switzerland. Recognized as a Top Place to Work for multiple years, XYZ fosters a dynamic and inclusive workplace. The company is committed to gender and diversity strategies, actively working to increase female and minority representation within its teams. XYZ prioritizes the well-being of its employees, offering optimal life-work balance through part-time and flex-time work options.

Values and Policies:

- **Gender & Diversity Strategy:** XYZ is dedicated to creating an inclusive workplace, implementing concrete actions to enhance female and minority representation in its teams.
- **Life-Work Balance:** The company actively supports an optimal life-work balance, providing part-time and flex-time work options to its employees.
- **Onboarding Program:** XYZ offers a complete and structured onboarding program, combining e-learning and on-the-job training to familiarize employees with the company, their role, and anticipated challenges.
- **Further Education Opportunities:** After onboarding, employees have access to additional education and training opportunities, contributing to continuous professional development.





Candidates Profiles:

Anna (they/their)

Curriculum Vitae:

Personal Information:

- Name: Anna
- Date of Birth: January 18, 2000
- Nationality: Italian
- Marital Status: Single
- Gender: not-specified
- Location: Milan, Italy
- Visa: not available

Education:

MA Student in Industrial Engineering

- University of Milan, Expected Graduation Year: 2024

Bachelor Degree in Engineering

- University of Milan, Graduation Year: 2020

High School Diploma in Linguistic/Foreign Languages

- Linguistic High School, Graduation Year: 2016

Work Experience:

Retail Sector

- Store Manager
- Employer: FashionEmporium, Milan
- Duration: 3 years (2018-2021)
- Responsibilities:
 - Managed day-to-day operations as a store manager.
 - Oversaw staff scheduling, training, and performance.
 - Maintained inventory and ensured visual merchandising standards.
 - Provided excellent customer service and contributed to sales growth.

Translation Internship

- Intern Junior Translator
- Employer: Multilingual Solutions, Rome
- Duration: 1 year (2017-2018)
- Responsibilities:
 - Assisted in translation tasks and language-related projects.
 - Collaborated with senior translators to ensure accurate and effective communication.

Skills:

- Team-Working Attitude
- Computer Skills (MS Office, Translation Software)
- Retail Management
- Visual Merchandising
- Customer Service

Languages:

- Italian: Native
- English: Fluent
- Spanish: Proficient

Hobbies and Personal Attitudes:

- Basketball
- Food Influencer on Instagram
- Team-Working Attitude
- Computer Skills



Anna (they/them)

Cover Letter:

Dear Hiring Manager,

I am writing to express my interest in the Assistant Store Manager/Salesperson position at XYZ. As a highly motivated individual with a background in retail management and translation, I believe my diverse skill set aligns well with the needs of your dynamic team.

During my three years as a Store Manager at FashionEmporium in Milan, I honed my leadership skills, oversaw day-to-day operations, and played a pivotal role in enhancing the overall customer experience. My background as an Intern Junior Translator at Multilingual Solutions in Rome allowed me to develop strong communication and language-related skills.

Currently pursuing a Master's in Industrial Engineering, with a Bachelor's Degree in Engineering and a High School Diploma in Linguistic/Foreign Languages, I bring a unique blend of technical and language proficiency to this role.

My passion for basketball and my role as a food influencer on Instagram demonstrate my commitment to teamwork, creativity, and effective communication. Combined with my computer skills, this diverse background positions me as an ideal candidate for this position. Due to my strong dedication to basketball, I am unavailable for work on Sundays as I typically have a scheduled match; I would make myself available on Saturdays and holidays.

I am deeply involved in LGBTQ+ civil rights, and, being a proud transgender woman myself, I actively promote an equal and inclusive society that embraces diversity and fosters understanding, acceptance, and respect for all individuals, regardless of their gender identity or sexual orientation.

Thank you for considering my application. I am eager to bring my energy and skills to XYZ and contribute to the continued growth of your flagship franchise.

Sincerely,

Anna





Diego (they/them)

Curriculum Vitae:**Personal Information:**

- Name: Diego
- Age: 30
- Nationality: Colombian
- Gender: non-binary
- Marital Status: not specified
- Location: London, UK
- Visa: available
- Disability Status: mild cerebral palsy

**Education:**

Bachelor University Degree:

- Major in Architecture, Universidad de los Andes - Bogotá (2015-2022)

High School Diploma

- Science High School of Bogotá (2015)

Work Experience:

Sport Retail

- Sales Clerk
- Bogotá, Colombia
- Duration: 2015-2019
- Responsibilities:
 - Took care of the company's orders and relationships with suppliers and customers
 - Monitored the phase of negotiations
 - Post-sales assistance management

Commercial Sector

- Clerk
- Bogotá, Colombia
- Duration: 2019-2021
- Responsibilities:
 - Reception service employee
 - Management of telephone calls at the reception
 - Customer assistance during negotiations and after-sales
 - Elaboration of estimates, commercial proposals, contract signing
 - Choice and organization of promotional material
 - Development of season ticket sales techniques

Store Assistant

- Employer: Retail Forever spa
- Duration: 2021-current
- Responsibilities:
 - Customer service
 - Sales support
 - Store cleaning and maintenance
 - Inventory management
 - Cash Handling
 - Product knowledge

Skills:

- Customer service
- Profitable management of working times
- Great organizational skills
- Purposeful and constructive attitude for achieving goals
- Architecture and Design
- Communication and teamwork

Languages:

- Spanish: Native
- English: Intermediate
- Italian: Intermediate

Hobbies:

- Hiking
- Scuba Diving (PADI license)
- Volleyball
- Dancing
- Ballet and Theatre

Diego (they/them)

Cover Letter:

Dear Hiring Manager,

I am writing to be considered in the Assistant Store Manager position at XYZ. I am 30 years old and have recently graduated in Architecture from the Bogotá Universidad de los Andes.



During my studies I have always worked, learning to manage stress and be flexible. I am temperamentally inclined to develop good relationships and to always maintain a positive approach towards the customer, I ensure maximum commitment and availability. I consider myself a sociable, dynamic, willing and eager to learn person.

I am also a sporty person. I have been playing volleyball at a competitive level for 20 years, and it is precisely thanks to this sport that I have learned the spirit of sacrifice, team spirit and values such as knowing how to collaborate within a group to achieve a common goal.

My experience as a store assistant in a retail store gave me the opportunity to develop a versatile skill set encompassing customer service, sales support, and merchandising, enabling me to thrive in a dynamic and fast-paced environment.

At the moment I am looking for a new professional opportunity that will allow me to get to know a new, stimulating and constructive reality for my life path, and I really think that I would fit perfectly in a dynamic and fresh reality as XYZ. I am also available to work flexible hours and during weekends.

My architectural background allowed me to learn how to look at things from a different perspective, and I was able to transfer that knowledge in my current position: I am always able to find alternative solutions to unexpected problems.

I am excited about the opportunity to contribute to XYZ and believe that my skills in sales and architecture make me a strong fit for this role. Thank you for considering my application.

I look forward to the possibility of contributing to the continued success of XYZ.

Sincerely,
Diego





Felix (he/his)

Curriculum Vitae:**Personal Information:**

- Name: Felix
- Page: 22
- Nationality: Swiss
- Gender: Male
- Marital Status: Single
- Location: Olten, Switzerland
- Visa: available

Education:

University Dropout

- University of Zurich, Major: Political Science - didn't complete

High School Diploma

Work Experience:

Hospitality Sector

- Waiter/Bartender
- Various Restaurants, Switzerland
- Duration: Various short-term positions (2019-2022)
- Responsibilities:
 - Took orders and served food and beverages.
 - Maintained cleanliness and orderliness of the bar and restaurant.
 - Interacted with customers to ensure a positive dining experience.

Summer Jobs

- Lifeguard
- Local Swimming Pool, Switzerland
- Duration: Summer seasons (2018-2020)
- Responsibilities:
 - Ensured the safety of swimmers.
 - Conducted water rescues and first aid when necessary.

Sales Assistant

- Employer: Penny Supermarket
- Duration: 2 year
- Responsibilities:
 - Greeted and assisted customers in a friendly and helpful manner.
 - Customer assistance and support
 - Managed day-to-day store operations
 - Inventory control: restock, counts, discrepancies.

Skills:

- Hospitality and Customer Service
- Physical Fitness (Gym and Boxing)
- Adaptability
- Goal-oriented
- Available on weekends and holidays

Languages:

- German: Native
- English: Basic

Hobbies and Lifestyle:

- Gym and Boxing Enthusiast
- Partying and Socializing
- Enjoys a Nightlife Lifestyle
- Casual Interest in Tattoo Artistry
- Fashion and lifestyle
- Instagram influencer



Felix (he/his)

Cover Letter:

Dear Hiring Manager,

I am writing to express my interest in the Salesperson position at XYZ. My diverse background in hospitality and summer jobs has equipped me with some skills in customer service and adaptability.

While I didn't complete university, I've gained practical experience working as a waiter, bartender, and lifeguard. These roles have developed my communication skills and ability to work in fast-paced environments. Although I briefly pursued tattoo artistry, I realized drawing isn't my forte. Despite these experiences, I believe I can bring a unique perspective and energy to your team.

I'm great with people, and if I see a benefit, I can persuade anyone to make a purchase. I learned these skills by socializing at parties and spending time with friends.

My passion for physical fitness, including regular gym sessions and boxing, reflects my dedication to staying physically active and maintaining a healthy lifestyle. I thrive in social settings, enjoying the energy of parties and nightlife.

While I may not have direct experience as a salesperson, I am motivated to take on this role to support my lifestyle and contribute to XYZ's team. I am confident that my commitment can make a positive impact.

I had some experience as a sales assistant, so I know I would bring a customer-focused approach, effective communication skills, and a proven ability to contribute to sales success in a retail setting.

Thank you for considering my application. I am eager to discuss how my skills align with your team's needs.

Sincerely,
Felix





Mark (he/his)

Curriculum Vitae:**Personal Information:**

- Name: Mark
- Date of Birth: September 1, 1995
- Marital Status: Single
- Location: Geneva, Switzerland
- Visa: available
- Disability Status: attention-deficit/hyperactivity disorder

**Education:**

Bachelor Degree in Economics

- University of Munich, Graduation Year: 2017

Technical Institute Diploma

- Technical Institute of Technology, Graduation Year: 2015

Work Experience:

Retail Sector

- Assistant Store Manager and Logistics Officer
- Employer: SuperMarket Retail, Munich
- Duration: 2 years
- Responsibilities:
 - Collaborated with the store manager for commercial and economic management.
 - Managed personnel, determined work shifts, and organized training activities.
 - Oversaw logistics operations to ensure smooth supply chain management.
 - Maintained store appearance and ensured legal compliance.
 - Provided excellent customer assistance and acted as the store manager when needed.

Production Quality Control

- Employer: TechPro Manufacturing, Berlin
- Duration: 3 years
- Responsibilities:
 - Conducted quality control checks in a production environment.
 - Ensured products met specified quality standards.
 - Collaborated with production teams to implement quality improvements.
 - Maintained detailed quality control records.

Skills:

- Leadership Skills
- Empathy and Attention Towards the Customer
- Multilingual: German (Native), English (Fluent), French (Proficient)
- Yoga and Meditation Teacher

Languages:

- German: Native
- English: Fluent
- French: Proficient

Hobbies and Personal Attitudes:

- Yoga and Meditation
- Leadership Skills Development
- Customer-Centric Approach
- Continuous Learning and Development

Mark (he/his)

Cover Letter:

Dear Hiring Manager,



I am excited to apply for the Assistant Store Manager/Salesperson position at XYZ. With a Bachelor's Degree in Economics and a robust background in both retail and production quality control, I am confident in my ability to make a meaningful contribution to your team.

During my tenure at SuperMart Retail in Munich, where I served as an Assistant Store Manager and Logistics Officer for two years, I honed my skills in personnel management, training coordination, and logistics oversight. My proactive approach ensured the store's compliance with legal standards and contributed to a positive customer experience. Additionally, my role allowed me to step in as the acting store manager when necessary, showcasing my versatility.

In my three years at TechPro Manufacturing in Berlin, I strengthened my commitment to quality control in a production setting. I maintained meticulous records, conducted thorough quality checks, and collaborated with production teams to implement improvements. This experience heightened my attention to detail and commitment to excellence.

Beyond my professional experiences, I am a dedicated yoga and meditation teacher. This passion has not only enhanced my leadership and communication skills but has also instilled in me a deep sense of empathy and attentiveness towards customer needs. Being multilingual—fluent in German and English and proficient in French—I bring a global perspective to my interactions.

Since I was little, I have loved animals and open-air activities: I am a mountain enthusiast and I love hiking with my friends and family. I try to spend as much time as possible in the mountains, hiking, climbing and skiing, so on the weekends I'm usually unavailable (no wifi on the mountain crest!).

Thank you for considering my application. I am enthusiastic about the opportunity to bring my skills, passion, and dedication to XYZ, and I look forward to the possibility of discussing how my experiences align with your team's needs.

Sincerely,

Mark





Mohammed (he/his)

Curriculum Vitae:

Personal Information:

- Name: Mohammed
- Age: 40
- Nationality: Moroccan
- Gender: Male
- Marital Status: Single (with children)
- Location: Marrakech, Morocco
- Visa: in process
- Disability Status: Hearing impairments

Education:

- High School Diploma

Work Experience:

Retail Sector

- Regional Manager
- Employer: MegaMart Retail, Morocco
- Duration: 20+ years (1998-2021)
- Responsibilities:
 - Started in entry-level positions and progressively moved up the ranks.
 - Managed stores of varying sizes, from small boutiques to large retail chains.
 - Promoted to Regional Manager, overseeing a chain of stores across the region.
 - Demonstrated creativity in visual merchandising and store presentation.

Boutique Fashion

- Store Manager
- Employer: ChicCloset Boutique, Casablanca
- Duration: 5 years (1993-1998)
- Responsibilities:
 - Managed day-to-day operations of a boutique fashion store.
 - Implemented effective visual merchandising strategies.
 - Led and motivated a team to achieve sales targets.

Electronics Retail

- Assistant Manager
- Employer: Tech Haven Electronics, Rabat
- Duration: 3 years (1990-1993)
- Responsibilities:
 - Assisted in managing store operations and staff.
 - Contributed to the implementation of effective sales strategies.
 - Ensured a positive customer experience.

Home Furnishings

- Department Manager
- Employer: CasaVista Furnishings, Marrakech
- Duration: 4 years (1986-1990)
- Responsibilities:
 - Managed the home furnishings department within a large retail store.
 - Oversaw inventory, pricing, and customer service.
 - Implemented merchandising strategies for increased sales.

Supermarket Chain

- Shift Supervisor
- Employer: GreenGrocers Supermarket, Tangier
- Duration: 4 years (1982-1986)
- Responsibilities:
 - Supervised daily shift operations.
 - Managed staff scheduling and training.
 - Ensured a well-maintained and customer-friendly store environment.

Clothing Retail

- Sales Associate
- Employer: Style Wear Apparel, Agadir
- Duration: 2 years (1980-1982)
- Responsibilities:
 - Started as a Sales Associate, assisting customers and maintaining store appearance.
 - Developed strong product knowledge and customer service skills.

Skills:

- Extensive Retail Management Experience
- Leadership and Regional Management
- Visual Merchandising and Store Presentation
- Fluent in Multiple Languages (Arabic, English, German)
- Creative Thinking
- Charismatic Leadership

Languages:

- Arabic: Native
- English: Fluent
- German: Basic

Hobbies and Personal Attitudes:

- Painting (Side Job)
- Squash and Golf
- Charismatic and Motivational Leadership
- Adaptability and Resilience



Mohammed (he/his)

Cover Letter:

Dear Hiring Manager,

I am writing to express my sincere interest in the Salesperson position at XYZ. With over 20 years of diverse retail management experience, I bring a unique blend of leadership, creativity, and linguistic proficiency to your esteemed team.

I hail from a highly impoverished village in central Morocco, and my journey to Switzerland was prolonged and marked by considerable hardship. I am familiar with the challenges of exerting considerable effort yet being unable to adequately provide for my family. However, my persistent desire to rise above challenging circumstances has been a driving force in my life. My story serves as a testament to the transformative power of a determined commitment to change. It illustrates that with a willingness to effect positive change, one can truly overcome substantial obstacles.

My journey in the retail sector, starting from the entry levels and growing to the role of Regional Manager in Morocco, has equipped me with a deep understanding of store operations, visual merchandising, and effective leadership. I am fluent in Arabic, English, and German, facilitating effective communication with a diverse clientele.

In addition to my extensive retail experience, I am a passionate painter, demonstrating my creative thinking and ability to bring a unique perspective to the sales environment. Known for my charismatic leadership style, I have successfully motivated teams and cultivated a positive and productive work atmosphere. I love spending time with my family and that is how I usually like to spend my weekends. If it is possible, I would prefer keeping this habit as much as possible in my work.

Living near Olten, Switzerland, I am excited about the opportunity to bring my skills and experience to XYZ, a company renowned in the large-scale retail trade. My proficiency in languages, creative thinking, and adaptability make me a strong candidate for this position.

Thank you for considering my application. I am eager to contribute to the success of XYZ and look forward to the possibility of discussing how my skills align with your team's needs.

Sincerely,
Mohammed





Priya (she/her)

Curriculum Vitae:**Personal Information:**

- Name: Priya
- Date of Birth: May 3, 1990
- Location: Lausanne, Switzerland
- Nationality: Indian
- Visa: available
- Marital Status: Married, 6 Children

Education:

High School Diploma in Industry-Technical Institute

- Lausanne Technical Institute, Graduation Year: 2008

Work Experience:

Food Manufacturing

- Quality check and process control
- Employer: Swiss Taste Manufacturing, Lausanne
- Duration: 5 years (2013-2018)
- Responsibilities:
 - Managed various aspects of food manufacturing processes.
 - Ensured adherence to quality and safety standards.
 - Collaborated with cross-functional teams for efficient production.

Retail Sector

- Customer Service and assistance
- Employer: Lausanne Retail Hub
- Duration: 2 years
- Responsibilities:
 - Provided excellent customer service and maintained store appearance.
 - Collaborated with the team to optimize store operations.

Customer Support Representative

- Employer: Vodafone
- Duration: 1 year
- Responsibilities:
 - Provided excellent customer service through phone support in a fast-paced call center environment.
 - Assisted customers with inquiries, resolved issues, and ensured satisfaction with products/services.
 - Collaborated with team members and other departments to address customer needs effectively.

Store Assistant

- Employer: Lidl Supermarket
- Duration: 1 year
- Responsibilities:
 - Managed day-to-day store operations and supported the overall success of the retail establishment.
 - Developed and implemented efficient inventory management procedures.

Skills:

- Problem-Solving and Team-Working Attitude
- Microsoft Office Knowledge
- Experience in Gaming

Languages:

- Native: Indian
- Proficient: English

Hobbies and Personal Attitudes:

- Gaming
- Problem Solving
- Team-Working Attitude

Priya (she/her)

**Cover Letter:**

Dear Hiring Manager,

I am writing to express my genuine enthusiasm for the Assistant Store Manager/Salesperson position at XYZ. With a High School diploma in Industry-Technical Institute and a rich background in both food manufacturing and the retail sector, I am eager to contribute my diverse skill set to your esteemed team.

My three-year tenure at Swiss Taste Manufacturing provided me with valuable insights into food manufacturing processes, emphasizing the importance of quality and safety standards.

What sets me apart is my proficiency in Microsoft Office, problem-solving attitude, and strong team-working skills. My passion for gaming has honed my multitasking and time management abilities, making me adept at handling complex situations with ease.

I am drawn to XYZ's commitment to excellence and innovation in large-scale retail trade. Your emphasis on creating an inclusive workplace aligns with my values, and I am excited about the prospect of contributing to XYZ's continued success.

I am deeply passionate about ecological activism, actively participating as a member of the Extinction Rebellion. Recently appointed as the director overseeing the organization of strikes and sit-ins, I faced a brief period of incarceration in 2020 lasting for two days due to my commitment to the cause.

I am an optimistic person with a rational approach towards life. I am committed to my work and career and do not hesitate in doing hard work. Able to coordinate various activities and have immense desire to learn new skills to improve performance. Unfortunately, I wouldn't be able to give my full availability to weekends and holidays, as I usually like to reserve these times to the ecological cause.

Thank you for considering my application. I look forward to the possibility of discussing how my skills align with your team's needs in more detail.

Sincerely,
Priya





29. Annex 15 – The Candidate Ranker: guidelines and instructions

Additional instructions sent via email before the workshop:

List of Features:

- Perfect match (100/100) with must-have requirements
- Perfect match (100/100) with nice-to-have requirements
- Education in relevant field/discipline
- Formal education level
- Relevant professional experiences
- Having professional experiences in a different field of interest
- Quantity of previous experiences (the more the better)
- Duration of previous experiences (the longer the better)
- Relevance ONLY of the latest work experience
- Prestige of past experiences
- Meeting ONLY the hard skills
- Language Level requested in the job offer
- Having the required licenses
- Candidate matching Diversity & Inclusion policies' criteria
- Cultural and Value fit for the company
- Availability to flexible working hours
- Distance from candidate and job's location
- Not being overqualified
- Law and policy compliance (visa requirements etc)
- Communication and social skills
- Different interesting hobbies
- Motivation and Resourcefulness
- Lack of experience compensated by a strong educational background

Questions:

Please take a moment to reflect and document (using your preferred format) the following points:

1. Which are the main features of the ones listed above that you consider relevant to select the right candidate for this position?
2. How would you rate each feature between “must-have”/”nice-to-have”/”irrelevant” ?
3. Are there any factors/features that you can think of that are not listed but are relevant to the selection of the fittest candidate?

Technical instructions to use the Candidate Ranker:

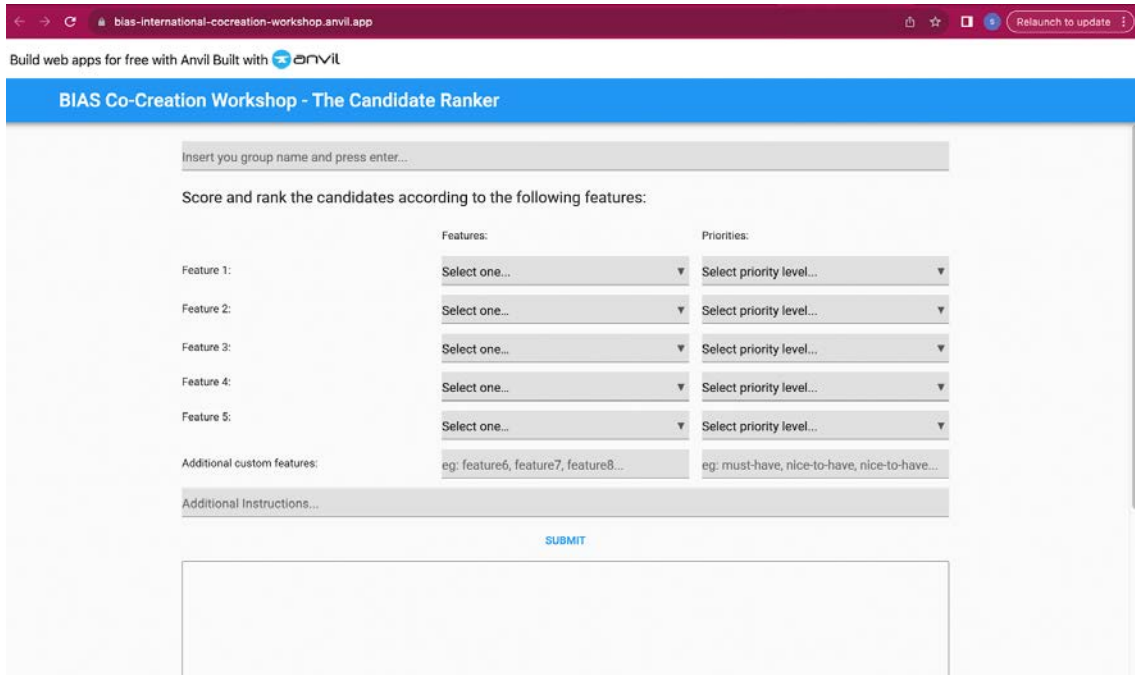
1. Open the link:

<https://bias-international-cocreation-workshop.anvil.app>

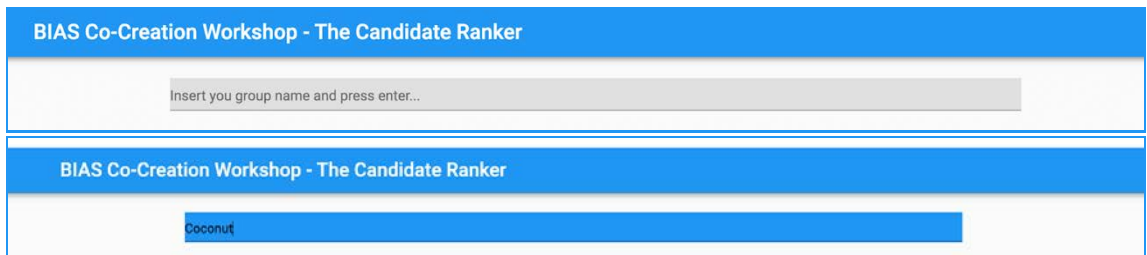




2. You will land on the following page:



3. The first thing to do before anything else is to insert your group name in the text box at the top of the page, and press Enter, in this way the system will be able to record your data. Once you press Enter, the box will turn blue to confirm that you have been registered.



4. The tool is able to extract the *must-have* and the *nice-to-have* requirements of a job ad and use them to rank the candidates based on their matching level with respect to the requirements; on top of this, the tool is configurable, meaning that you can set the features over which it must evaluate the candidates: you can select the features from a pre-set list (the list we sent you in our email) that will appear as a drop-down menu when clicking on “*Select one...*”, and also select the level of importance that you want to assign to each selected feature (is it a *must-have* or a *nice-to-have*?, which is also a drop-down box). For example, you can select Feature 1 as “*Formal education level*” and set the priority for this feature to “*must-have*”. It means that the tool will consider this factor with high importance and will assign more value to the candidates satisfying this requirement in the final ranking.



BIAS Co-Creation Workshop - The Candidate Ranker

Coconut

Score and rank the candidates according to the following features:

	Features:	Priorities:
Feature 1:	Formal education level	Select priority level... Must-have
Feature 2:	Quantity of previous experiences (the m	✓ Nice-to-have
Feature 3:	Select one...	Select priority level...
Feature 4:	Select one...	Select priority level...
Feature 5:	Select one...	Select priority level...
Additional custom features:	eg: feature6, feature7, feature8...	eg: must-have, nice-to-have, nice-to-have...

Additional Instructions...

SUBMIT

BIAS Co-Creation Workshop - The Candidate Ranker

Coconut

Score and rank the candidates acco

	Features:	Priorities:
Feature 1:	Select one...	
Feature 2:	Quantity of previous experiences (the more the better)	
Feature 3:	Relevance ONLY of the latest work experience	
Feature 4:	Meeting ONLY the hard skills	
Feature 5:	Language Level requested in the job offer	
Additional custom features:	eg: feature6, feature7, feature8...	ice-to-have...

Additional Instructions...

In the page you see 5 features to select but please consider:

1. Not all the 5 fields need to be filled. You can select even just one feature.
2. If you set a feature you also need to specify its corresponding level of importance.
3. If you want to add more features (or you just want to write your own features) you can do it by manually writing them in the *additional custom features*: they just need to be separated by a comma.
4. For every additional feature in the *additional custom features* you have to manually write its level of priority (*must-have* or *nice-to-have*), also in a form of list separated by commas. For example, if you want to specify the additional features “creativity” as a





must-have (or just *must*) and “*not being under qualified*” as a nice-to-have requirement (or just *nice*), you have to write under Features “*creativity, not being under qualified*” and under Priorities “*must, nice*”.

BIAS Co-Creation Workshop - The Candidate Ranker

Coconut

Score and rank the candidates according to the following features:

	Features:	Priorities:
Feature 1:	Formal education level ▼	Must-have ▼
Feature 2:	Quantity of previous experiences (the m ▼	Nice-to-have ▼
Feature 3:	Select one... ▼	Select priority level... ▼
Feature 4:	Select one... ▼	Select priority level... ▼
Feature 5:	Select one... ▼	Select priority level... ▼
Additional custom features:	Creativity, not being under qualified	must, nice
Additional Instructions...		
SUBMIT		

5. Upon clicking the SUBMIT button, you transmit your configuration to the system. There's no requirement to upload texts related to the job ad and candidates, as they are already stored in the system.

6. The output will be displayed in the large text box area at the bottom of the page, so please do not write anything there. The format of the output would be as follows: the system presents each candidate in order, showing their ranking position, their total suitability score on a scale of 100, and a brief explanation of the ranking and scoring choices.

```
Candidate:
Ranking:
Explanation:
Total Suitability Score:
```

7. Above the SUBMIT button, there is an Additional Instructions box available for use, if necessary, to specify requests regarding the desired output format. For instance, if you wish the system to assign a score to each feature you've set, you can articulate this request in the box, such as: "Assign a score on a scale of 100 to every candidate for every feature." Whether you need more detailed explanations or a more schematic presentation, you can include these instructions in the box. However, exercise caution in utilizing this functionality, as the system operates on a pre-written prompt with specific instructions. Adding too many extra instructions may risk confusing the system and potentially leading to hallucinations.





BIAS Co-Creation Workshop - The Candidate Ranker

Feature 5:

Additional custom features:

Additional Instructions...

[SUBMIT](#)

Candidate: Anna
 Ranking: 1
 Explanation: Anna holds a Bachelor's degree in Engineering, which shows a high level of formal education. In addition, she is currently pursuing a Master's degree in Industrial Engineering, displaying her dedication to continuous learning. Anna has extensive retail management experience, having served as a store manager at FashionEmporium where she oversaw staff scheduling, training, and performance, and maintained inventory and visual merchandising standards. She demonstrates creativity through her role as a food influencer on Instagram. However, she won't be able to work on Sundays.
 Total Suitability Score: 90

Candidate: Mohammed
 Ranking: 2
 Explanation: Mohammed started his career in retail at an entry-level and has now risen to the position of Regional Manager. This showcases not only his extensive experience in retail but also his commitment towards his career. Although he does not hold a formal degree, he has demonstrated significant progress in his career. Mohammed also shows creativity in his merchandising strategies as well as painting. However, he prefers spending his weekends with his family, which might compromise his job availability during these days.
 Total Suitability Score: 85

Candidate: Prya
 Ranking: 3

8. You can change the configurations as many times as you want and click *SUBMIT* to regenerate the outcome. The important aspect is that if you select a feature, you also select its level of importance, otherwise the system will warn you with an error message. Please also make sure to use commas to separate the manually written features one from the other.



Guidelines to participants and facilitators:

1. **Individual reading of profiles focusing on strengths/weaknesses and potential bias (10 min).** Within the printed documents, you have the job advertisement received via email, along with a list of features and questions. Please take a moment to individually reflect on the provided questions. Additionally, there are six distinct candidate profiles, each accompanied by their CV and cover letter. Distribute the reading among team members to ensure that each profile is reviewed, with each team member responsible for one profile while collectively covering all six. (10 min). While reading the profiles, please individually highlight on the text two things:
 - a. The elements of strengths and weaknesses of the candidate, with respect to the requirements of the job ad.
 - b. The elements of potential biases in the CV and cover letter.
2. **Understanding the Candidate Ranker tool (5 min).** Each group's facilitator will now briefly explain how the Candidate Ranker tool works, there are also printed instructions for the participants to read. (5 min)
3. **Prioritizing features and configuring the tool (10 min).** Exchange information about each candidate with one another to collaboratively arrive at a consensus regarding the features and priorities for configuring the tool. If you have conflicting opinions do not worry, you can configure the system more than once, so feel free to "try and see". Please select:
 - a. Which features to set (from the pre-set list or not)
 - b. Which priority to give to every feature
 - c. Any additional instructions to give the tool
4. **Discuss on the ranking and explanations (20 min).** Once you have inserted all the inputs you prefer, click on *SUBMIT* and reflect on the outcome. Take a moment to look at the following questions and discuss together:
 - a. Do you agree with the final ranking?
 - b. Are the explanations sufficient for decision-support?
 - c. What does this version of the tool do well? Are there any desired missing functionalities?
 - d. Try to change the configuration: select other features, change the priority level, invent new features that you think are appropriate:
 - i. Did you see any differences in the output?
 - ii. Did you expect these changes? Do you agree with them?
5. **Discussion in view of the plenary session (15 min):** reflect on the overall exercise. What are your perceptions of the tool? Particularly, try to cover the following points:
 - a. What are the potential hazards of using such a tool?
 - b. Would you trust this tool? If not, how would you change it to make it trustworthy?
 - c. Do you find this tool fair? If not, how would you change it to make it fairer?





30. Annex 16 – The Mitigation Tool: guidelines and instructions

Technical instructions to use the tool:

1. Open the link:

<https://bias-international-cocreation-workshop.anvil.app>

2. You will land on the following page:

3. The first thing to do before anything else is to insert your group name in the text box at the top of the page, and press Enter, in this way the system will be able to record your data. Once you press Enter, the box will turn blue to confirm that you have been registered.

4. If you're familiar with ChatGPT, it operates similarly: you input instructions in the "Write here..." text box, and the tool generates a response displayed in the *Output* text box.
5. Feel free to input instructions and click the ASK button as many times as you like. Keep in mind that your input will be incorporated into a broader prompt, encompassing all candidates' profiles and specifying the output format. For the purpose of this exercise, we'll provide a list of instructions to choose from, to avoid participants falling out of scope and help them navigate through the tool. All the instructions are all related to the identification of biases in the profiles.





BIAS Co-Creation Workshop - The Mitigation tool

Doconut

Add your instructions:

For every candidate, display all the possible negative biases that a recruiter can identify when deciding who is the fittest candidate.

ASK

Output

6. The output has the following format: for every candidate a list of identified elements is displayed, together with a list of mitigation actions suggested.

BIAS Co-Creation Workshop - The Mitigation tool

For every candidate, display all the possible negative biases that a recruiter can identify when deciding who is the fittest candidate.

ASK

Prya (she/her)

Potential Bias:

1. Prya has mentioned multiple times that she is not available to work during weekends or holidays (which is a requirement for the job) because of her ecological commitments. This might bias the recruiter.
2. Apart from customer service roles, Prya does not have direct experience in store management or other required roles, which might cause recruiters to disregard her.

Mitigation Suggestions:

1. Make sure to see Prya's commitment to ecological causes as evidence of passion and initiative, which could be beneficial to the company in terms of environmental responsibility and sustainability efforts. Negotiate the schedule issue during the interview process.
2. Look beyond her current job titles and instead focus on the relevant skills she has gained, such as customer service and adherence to quality and safety standards. She could be trained for the remaining skills once on the job.

Anna (they/them)

Potential Bias:

1. Anna's application comes from Italy, and her visa status is unavailable, which might bias the reviewer.
2. Though Anna has the necessary experience and skill set, she clarified that she can't work on Sundays due to her basketball matches.

Prompt to complete:

For every candidate identify all the elements of biases related to the following types:

- [type1]
- [type2]
- [type3]





Guidelines to participants and facilitators:

1. **Share on potential biases (10 min).** Briefly share among your group which can be the elements that you have identified as potentially subject to bias when reading their profiles. Reflect on intersected differences such as potential prejudices based on:

KEY BIASES

- age
- sex
- gender
- gender identity
- sexual orientation
- diverse abilities
- skin color
- race
- ethnicity
- religion
- social status
- economic background

Please consider both positive and negative types of biases that can originate from the above mentioned individual features.

Please select three types of bias from the list above that you consider as most relevant for the profiles.

Also, **consider additional types of bias** that are crosscutting and are more specific for HR practice/process from the list below, **and select one** that you would consider important to analyze.

ADDITIONAL BIASES

- a. **Conformity Biases:**
Tendency to align with prevailing opinions or behaviors, impacting independent judgment.
 - b. **Horn-Effect / Halo-Effect:**
Overemphasis on single negative (Horn) or positive (Halo) traits, influencing overall perception.
 - c. **Confirmation Biases:**
Inclination to favor information confirming pre-existing beliefs.
 - d. **Illusory Correlation Biases:**
Incorrectly perceiving a relationship between unrelated events or characteristics.
 - e. **Biases Related to the Specific job Position (Store Assistant Manager):**
Prejudices specific to the role, potentially affecting decision-making.
 - f. **Biases in AI Tool for Ranking Candidates:**
Unfair inclinations embedded in the algorithms, impacting candidate assessment.
2. **Learn how the tool works (5 min).** Each group's rapporteur will now briefly explain how The Mitigation tool works, there are also printed instructions for the participants to read.
 3. **Prompt the system (5 min)** Each group works on the 3 previously selected key bias + 1 additional bias. Use the following text structure and insert the 4 types of bias to complete the instructions. (5 min)
For every candidate identify all the elements of biases related to the following types:





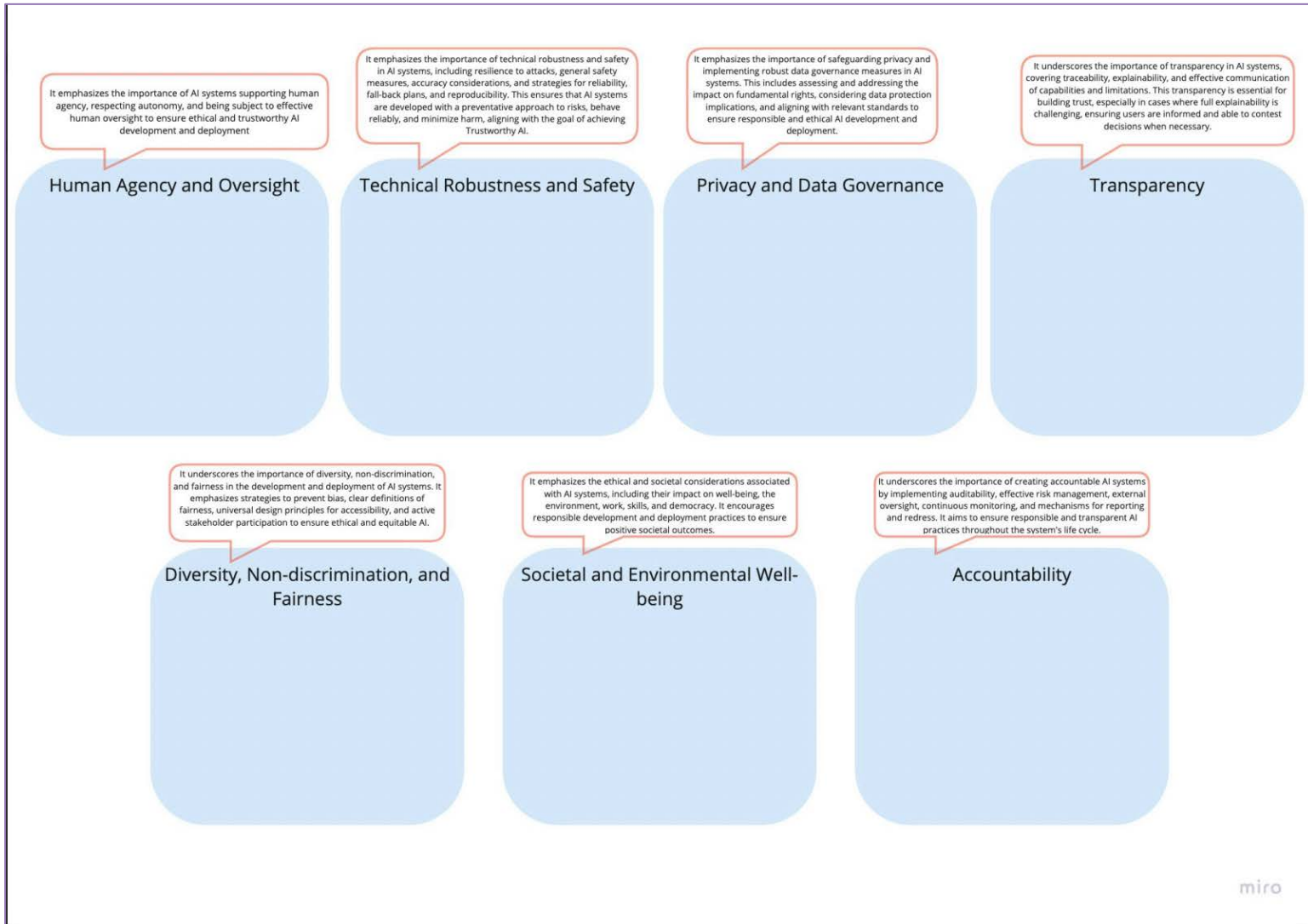
- [type1]
 - [type2]
 - [type3]
 - [type4]
4. **Review the output from the system (15 min).** Copy and paste the prompt with the inserted types of biases, then click on the *ASK* button and read the outcome, it should present the identified biases for each candidate, and also propose one or more mitigation strategies to avoid them. Try to reflect on the following points:
 - a. Reflect on the overall experience:
 - I. Were there any biases pointed out by the group that were not identified by the system?
 - II. Were there any biases pointed out by the system that were not identified by the group?
 - b. Are the explanations of the biases provided by the system reasonable?
 - c. What does this version of the tool do well? Are there any desired missing functionalities?
 5. **Discussion in view of the plenary session (15 min):** reflect on the overall exercise. What are your perceptions of the tool? Particularly, try to cover the following points:
 - a. What are the potential hazards of using such a tool?
 - b. Would you trust this tool? If not, how would you change it to make it trustworthy?
 - c. Do you find this tool fair? If not, how would you change it to make it fairer?





31. Annex 17 – The ALTAI requirements brainstorming

The poster:



miro





Guidelines for the exercise:

The Assessment List for Trustworthy Artificial Intelligence (ALTAI) is a self-evaluation tool developed by AI HLEG (the High-Level Expert Group on Artificial Intelligence). Following a piloting process where over 350 stakeholders participated, an earlier prototype of the list was revised and translated into a tool to support AI developers and users in developing Trustworthy AI.

The tool supports the actionability of the key requirements outlined by the Ethics Guidelines for Trustworthy Artificial Intelligence (AI), presented by the High-Level Expert Group on AI (AI HLEG) presented to the European Commission, in April 2019. The Ethics Guidelines introduced the concept of Trustworthy AI, based on seven key requirements:

1. Human Agency and Oversight

It emphasizes the importance of AI systems supporting human agency, respecting autonomy, and being subject to effective human oversight to ensure ethical and trustworthy AI development and deployment

2. Technical Robustness and Safety

It emphasizes the importance of technical robustness and safety in AI systems, including resilience to attacks, general safety measures, accuracy considerations, and strategies for reliability, fall-back plans, and reproducibility. This ensures that AI systems are developed with a preventative approach to risks, behave reliably, and minimize harm, aligning with the goal of achieving Trustworthy AI.

3. Privacy and Data Governance

It emphasizes the importance of safeguarding privacy and implementing robust data governance measures in AI systems. This includes assessing and addressing the impact on fundamental rights, considering data protection implications, and aligning with relevant standards to ensure responsible and ethical AI development and deployment.

4. Transparency

It underscores the importance of transparency in AI systems, covering traceability, explainability, and effective communication of capabilities and limitations. This transparency is essential for building trust, especially in cases where full explainability is challenging, ensuring users are informed and able to contest decisions when necessary.

5. Diversity, Non-discrimination, and Fairness

It underscores the importance of diversity, non-discrimination, and fairness in the development and deployment of AI systems. It emphasizes strategies to prevent bias, clear definitions of fairness, universal design principles for accessibility, and active stakeholder participation to ensure ethical and equitable AI.

6. Societal and Environmental Well-being

It emphasizes the ethical and societal considerations associated with AI systems, including their impact on well-being, the environment, work, skills, and democracy. It encourages responsible development and deployment practices to ensure positive societal outcomes.

7. Accountability

It underscores the importance of creating accountable AI systems by implementing auditability, effective risk management, external oversight, continuous monitoring, and mechanisms for reporting and redress. It aims to ensure responsible and transparent AI practices throughout the system's life cycle.

Evaluating the Debiasser's trustworthiness:





Contemplate your morning interactions with the tools and explore how a bias-free AI system, referred to as the Debiaser, can fulfill the specified requirements while aiding decision-making in recruitment. Please elaborate on the following points and document your responses on post-its, using color-mapping logic. If needed, take a moment to discuss with your group over the interpretation of the requirements. Place each post-it in the designated box on the poster that aligns with its relevance:

1. Identify the most critical potential issues concerning the Debiaser for each listed requirement. **[red]**
2. For each listed requirement, outline solutions that can be taken in developing the system that can ensure Debiaser compliance. **[green]**
3. For each listed requirement, identify actions that can be taken by the recruiters (and all the stakeholders) that can ensure Debiaser compliance. **[blue]**
4. Develop metrics for measuring the fulfillment of each requirement by the Debiaser. **[violet]**

Notes Examples:

Guarantee data protection compliance	Guarantee a level of precision and out-of-the-box thinking of the entire process	Adaptive Learning: the tool should be able to receive feedbacks form recruiters to keep learning	User-friendly interface
Build an unsupervised-learning system that self-learns (no need for human intervention)	Build a human-IN-the-loop system (involve human intervention in every decision cycle)	Build a human-ON-the-loop system (allows human intervention during design and monitoring)	Build a human-in-command system (human oversees the system's overall activity)
Ensure consent from the candidate in handling their data	Avoid and mitigate subjectivity when screening a CV	Have countermeasures in place to mitigate occurring biases	The AI model needs to explain clearly and transparently the choices made
Ensure consistency (the same output for similar input/profiles)	The AI models should be validated in various use cases and applied only in the validated fields	Inform the candidate on the use of AI in the recruiting process	Undergoing periodic reviews where the data sources are inspected for any signs of bias or unfair weighting
Generate evaluations on multiple levels (multi-perspective)	Use a technology that avoids irrelevant elements which might generate bias	Monitor and measure performance to ensure the AI system is adding value to the process.	Periodical interviews with the users to asses the AI system
Have an AI ethic board to enforce responsible systems	Define objective criteria to evaluate the fairness of AI systems	Consider the hiring rate with respect to the employment duration (longer duration means better hiring)	The Debiaser could incorporate bias from historical data





32. Annex 18 – Learning needs

The brainstorming poster:



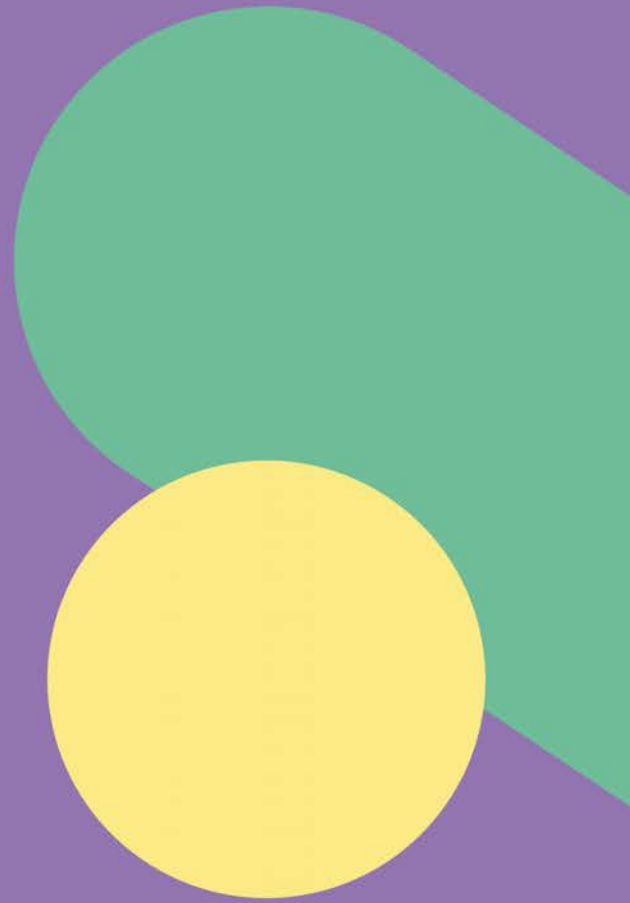
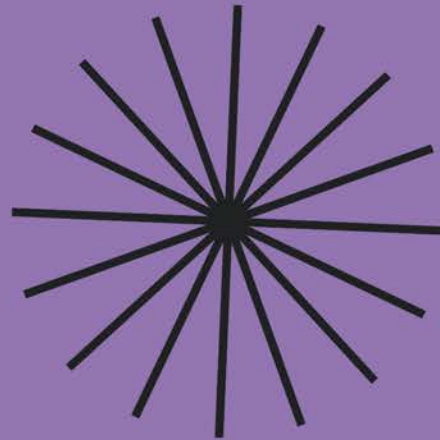
BRAINSTORMING ON LEARNING NEEDS

Existing courses/training on bias in AI or close-by topics	Knowledge/information you would expect to get from a a capacity building program
Skills that you would expect to develop/improve	Case studies to share (both from your organization and elsewhere)?
Ideal duration of a capacity building program	In person/online format?



BIAS

Mitigating biases
of AI in the
labour market



Consortium



NTNU



UNIVERSITY
OF ICELAND

LOBA*



CrowdHelix



SMARTVENCE



Universiteit
Leiden
eLaw



Digiotech

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](https://twitter.com/BIASProjectEU)

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