BIAS

Mitigating biases of AI in the labour market

D2.2
Preliminary report on co-creation methodologies and findings

Contact us

info@biasproject.eu

www.biasproject.eu

Funded by the European Union

@BIASProjectEU

BIAS

Mitigating biases of AI in the labour market

D2.2 Preliminary report on co-creation methodologies and findings

Deliverable Type

Report

Month and date of delivery

October 31., 2023

Work Package

2: Stakeholder involvement, needs assessment, co-creation

WP Leader

SVEN

Dissemination Level

Public

Deliverable Lead Beneficiary

SVEN

Programme	Contract Number	Duration	Start
HE	101070468	48 months	November 2023

Funded by the European Union

email
info@biasproject.eu

website
www.biasproject.eu

Main Authors

Name	Organisation
Marzia Cescon	SVEN, Smart Venice
Maria Sangiuliano	SVEN, Smart Venice

Contributors

Name	Organisation
Mascha Kurpicz-Briki and Alex Puttick	BFH, Berne University of Applied Sciences
Pinar Øzturk	NTNU, Norwegian University of Science and Technology

Peer Reviews

Name	Organisation
Öznur Karakas	NTNU, Norwegian University of Science and Technology
Carlotta Rigotti	ULEID, Leiden University

Revision History

Version	Date	Reviewer
0.1	22.09.2023	NTNU, ULEID
1.0 Final	30.10.2023	SVEN

Abbreviation	Meaning
AI	Artificial Intelligence
ATS	Applicant Tracking System
BS	Bachelor of Science
CBR	Case Based Reasoning
CSO	Civil Society Organization
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
SMEs	Small and Medium Enterprises
WEAT	Work Embedding Association Test
WP	Working package

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

1 Table of contents

1 Executive Summary 9

2 Introduction and methodological approach 10

3 Cross-cutting methodological aspects for the two workshops on engagement and preparation 14

3.1 Targeted participants/stakeholders 14

3.2 Preparatory and engagement activities 14

4 The first workshops' specific co-creation methodology 17

4.1 Workshop's agenda and target 17

4.2 Introduction & BIAS presentation 18

4.3 Panel discussion 18

4.4 Co-creation group work 20

4.5 Reporting process 25

5 Implementation of the first co-creation workshops: results 28

5.1 Overview of the conducted workshops 28

5.2 Overall implementation of the workshops 29

5.3 Summary of the reports 32

5.3.1 1st co-creation workshop in Italy 32
5.3.2 1st co-creation workshop in Norway 36
5.3.3 1st co-creation workshop in The Netherlands 40
5.3.4 1st co-creation workshop in Iceland 44
5.3.5 1st co-creation workshop in Türkiye 48
5.3.6 1st co-creation workshop in Estonia 52
5.3.7 1st co-creation workshop in Switzerland 56

5.4 Workshops' core outputs: the wordlists 60

5.4.1 Overview of the collected wordlists 60
5.4.2 Use of the wordlists in the BIAS technological development 63

6 Methodology of the second co-creation workshop 65

6.1 Workshop's agenda and target 65

6.2 Introduction & BIAS presentation 66

6.3 Discussion in two groups: how does a fair HR recruitment process look like? 66

6.4 Interactive/hands on work: which requirements for AI tools in recruiting? 69

6.5 Reporting process 76

7 Current state of the second co-creation workshops 77

8 Concluding remarks and next steps.....78

9 References.....80

Annex 1 – Facilitation principles & conflict management tips.....81

Annex 2 – Scenarios proposed.....85

List of tables

Table 1 Categories and numbers of stakeholders involved in the two workshops 15
Table 2 Categories and numbers of stakeholders involved in the first workshop 17
Table 3 PROs and CONs of the use of AI systems in recruitment and HR management 19
Table 4 Gender categories 21
Table 5 Structure and description of the exercise 24
Table 6 Examples of filling the spreadsheet 26
Table 7 Kind of report and related language 27
Table 8 Overview of participants in 1st co-creation workshops 28
Table 9 Categories and numbers of stakeholders at SVEN's first co-creation workshop 32
Table 10 Key takeaways from the Italian workshop 36
Table 11 Categories and numbers of stakeholders at NTNU's first co-creation workshop 37
Table 12 Key takeaways from the Norwegian workshop 40
Table 13 Categories and numbers of stakeholders at ULEID's first co-creation workshop 41
Table 14 Key takeaways from the Dutch workshop 44
Table 15 Categories and numbers of stakeholders at HI's first co-creation workshop 45
Table 16 Key takeaways from the Icelandic workshop 48
Table 17 Categories and numbers of stakeholders at FARPLAS' first co-creation workshop 48
Table 18 Key takeaways from the Turkish workshop 52
Table 19 Categories and numbers of stakeholders at DIGI's first co-creation workshop 53
Table 20 Key takeaways from the Estonian workshop 56
Table 21 Categories and numbers of stakeholders at BFH's first co-creation workshop 57
Table 22 Key takeaways from the Swiss workshop 60
Table 23 Other grounds of discrimination addressed by partners in the spreadsheets 61
Table 24 Categories and numbers of stakeholders involved in the second co-creation workshop 65
Table 25 Structure of the second workshop 66
Table 26 Scenario and persona of the second workshop 70
Table 27 Scenario and persona of the second workshop 74
Table 28 Structure and description of the second group work 75

List of figures

Figure 1 Stakeholders per category 29
Figure 2 Gender of participants 29
Figure 3 Main bias identified in the workshops 31
Figure 4 1st co-creation workshop at SVEN 32
Figure 5 1st co-creation workshop at NTNU 37
Figure 6 1st co-creation workshop at ULEID 41
Figure 7 1st co-creation workshop at FARPLAS 48
Figure 8 1st co-creation workshop at DIGI 52
Figure 9 Positive and negative bias associated to gender and race/ethnicity 61
Figure 10 Words/sentences per category 62
Figure 11 Words/sentences per category and kind of bias 63
Figure 12 Main points concerning fairness in HR recruitment/selection 67
Figure 13 Poster template for the second group work exercise 72

1 Executive Summary

This 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 Türkiye). 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) and the second round (in Chapter 6). 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. All internal templates, papers, and workshop material linked to in this deliverable will also be included as Annexes in the 1st Periodic Report. 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. Chapter 7 addresses overall concluding remarks and outlines the next steps of the BIAS co-creation process.

It is important to note that this report represents a preliminary and partial version of D2.4, the 'Final Report on Co-creation Methodologies and Findings,' which is due by M15 of the project.

2 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 Grant Agreement (GA), 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 by (Robertson et al., 2019):

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.

In the BIAS project's co-creation methodology, special attention was dedicated to thoroughly considering 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 experts in philosophy.

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 Debiaser and the fairness evaluation of both NLP and CBR-based tools.
- Identifying the ideal features of a Debiaser 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 Debiaser 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 Debiaser Tool and expanding understanding of fairness in HR concepts.

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 and 6). Both 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. Scenarios are regarded as a valuable tool in the relevant literature (Leikas at 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 both the first and second 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 Work 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 websites such as the Harvard project IMPLICIT[1]. 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. Based on the IAT, the Words Embeddings Association Test (WEAT) was developed to measure bias in word embeddings rather

[1] https://implicit.harvard.edu/implicit/

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 and philosophy experts, 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 endeavour, 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 cantered 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' technique[2] 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).

[2] https://www.mindtools.com/aiwjjpy/designing-future-state-customer-journeys

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

3.1 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 Debiaser 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 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 from June–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 €4 900 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.

3.2 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 identify 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. Some partners that did not have in-house expertise in facilitating participatory workshops (BHF, DIGI and FARPLAS) had dedicated resources (€7 000) for hiring facilitators.

SVEN 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.

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

- Thoroughly identify potential stakeholders belonging to the various categories to be involved, as listed in the table below:

Stakeholder type	**1st workshop**	**2nd workshop[3]**
HR officers and networks, associations of HR specialists	4	10–12
Workers and workers' representatives (e.g., trade unions)	8	2–4
AI specialist, practitioners, academics, researchers, students	4	4
Representatives of civil society organisations (e.g., associations, NGOs), networks, organisations advocating for equality	8	2–4
Legal experts in human rights and/or labour law		2
Experts in philosophy		2
Total:	24	24

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

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

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

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 and networks they already had in place in the frame of other projects or 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:

- 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.
- 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.

4 The first workshops' specific co-creation methodology

4.1 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 work sessions (2 hours and 15 min total including coffee break)
- Coffee/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 concerning discrimination grounds and axes of inequalities: gender and ethnic/cultural background. However, other potential biases were also explored. The first workshop involved the following stakeholder categories:

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

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

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.

4.2 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 BFH available at this link. 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 on next steps of BIAS:

- A **second co-creation workshop** scheduled to take place **from August–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.

4.3 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 already 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. Facilitators were advised to read an internal paper titled "An Introduction to the Fairness Notion for BIAS Project People," prepared by NTNU, (to be provided as an Annex in the 1st Periodic Report) in preparation for the meeting.

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.

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.

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

> **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 CONs?*

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.

4.4 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 **a comprehensive definition of "bias" was adopted**[4], 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. Likewise, 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 occupation[5], while "vertical segregation" pertains to the concentration of women and man in distinct grades, levels of responsibility or positions, as defined by EIGE[6]. 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 Report*[7] 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 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

[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

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.

Approaches to gender	**Persona 1**	**Persona 2**	**Notes**
Binary approach	Man	Woman	
Non-binary[8] approach	Cisgendered (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

Table 4 Gender categories

When referring to "non-cis persons", the intention was to include individuals who identify as queer, non-binary, or trans[9]. 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

[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.

Partners were asked to choose two race/ethnicity categories only to work on within the workshop: white (as the "majority ethnicity category" of the countries) 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 can be found this link. 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 four groups, six people per group: ideally one HR officer, one AI specialist, one workers/workers representatives/representatives of civil society organizations, NGOs, networks per group. Four rapporteurs had to be identified by partners and be in charge of notes-taking. The four rapporteurs could have been the two facilitators plus two additional people of the partner organization (e.g. students), or else four 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 found at this link, containing:

- The scenario/job offer selected (Selected from those presented in Annex 2
- Four "personas" profiles, completed as per described above (Selected from those presented in
- Four "personas" profiles, completed as per described above, but with no picture
- A blank page with space for participants to write a cover letter
- Template for the walking plenary session
- Reporting template for the note taker

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 be provided to the working groups.

The group work lasted around 2 hours and was structured as follows:

First activity - discussion on the job offer (15 min)	The HR officer's role in this part of the workshop is to go through the scenario/job offer and provide insights regarding: - 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. 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).
Second activity - elaboration of the cover letter (30 min)	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. 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. 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". 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.
15 minutes coffee break	
Third activity - discussion on the cover letter (20 min)	The group reads the cover letter and discusses with the HR officer on the following questions using sticky notes: - 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? 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.
Fourth activity - cover letter rephrasing (20 min)	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. 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.
Walking plenary (20 min)	Each group presents its result to the other groups using the posters and post-its, explaining: - 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.

Table 5 Structure and description of the exercise

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 at the following link. 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.

4.5 Reporting process

Partners had to report on the overall results of the workshops and especially of the group works using the templates included in this folder. The folder contains two templates:

- a document template for reporting the results of the workshop as a whole, 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 standards[10].
- 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, knowledge, and competencies beyond career-specific qualifications. The knowledge of languages was also considered and falls within this sub-category.

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

- 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:

	Career	Family
Black man		bambini (lavoro da remoto)
		children (remote working)

Personal info	
Personal attitudes	Hobbies/leisure
	yoga (gestione dello stress)
	yoga (stress management)

Table 6 Examples of filling the spreadsheet

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.

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

Table 7 Kind of report and related language

5 Implementation of the first co-creation workshops: results

5.1 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, out of the total KPI of 168 as seen in table 8 below:

Partner	Participants
Smart Venice	17
NTNU	23
Leiden University	20
University of Iceland	24
Farplas	18
Digiotouch	20
BFH	22
Total:	**144/168**

Table 8 Overview of participants in 1st co-creation workshops

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.

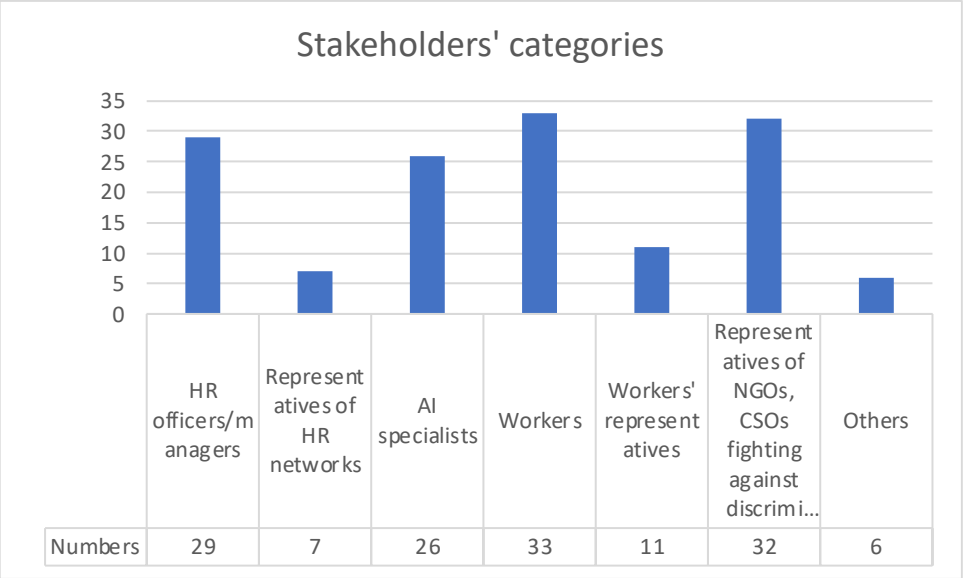

Figure 1 Stakeholders per category

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 the set KPI of minimum 45% included in the Education Action Plan 2021-2027.

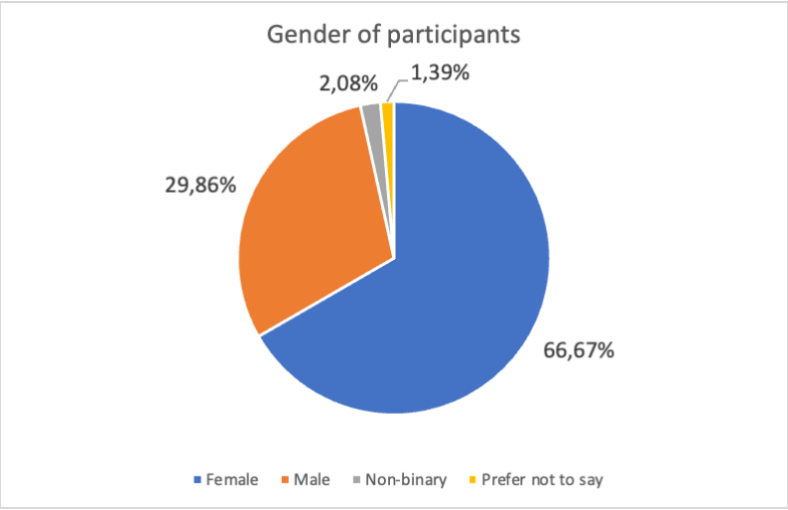

Figure 2 Gender of participants

5.2 Overall implementation of the workshops

The reports of the 7 workshops prepared by the involved BIAS partners can be found at this link. All of them 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 on 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).

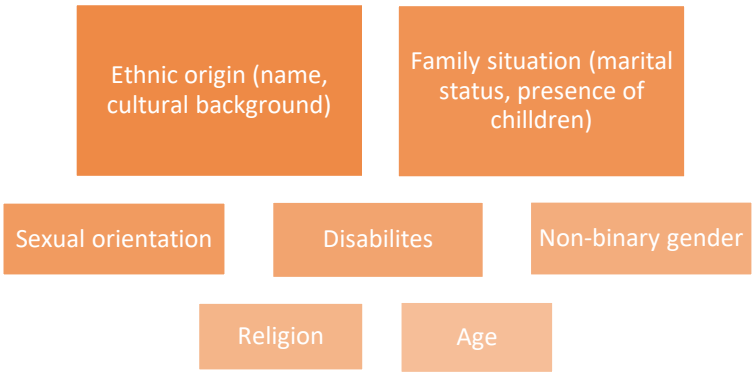

Figure 3 Main bias identified in the workshops

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.

5.3 Summary of the reports

5.3.1 1st co-creation workshop in Italy

Intro

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.

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
Total:	**17**

Table 9 Categories and numbers of stakeholders at SVEN's first co-creation workshop

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:

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.

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.

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 available here.

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.

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"

Table 10 Key takeaways from the Italian workshop

5.3.2 1st co-creation workshop in Norway

Intro

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:

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
Total:	**23**

Table 11 Categories and numbers of stakeholders at NTNU's first co-creation workshop

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:

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.

Question 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.

Question 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

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.

Question 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 available here.

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.

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	

Table 12 Key takeaways from the Norwegian workshop

5.3.3 1st co-creation workshop in The Netherlands

Intro

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:

Categories	Number
HR officers/managers	5
AI specialists	2
Workers	4
Representatives of NGOs, networks, organisations fighting against discriminations	6
Other	3
Total:	**20**

Table 13 Categories and numbers of stakeholders at ULEID's first co-creation workshop

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:

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.

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.

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.

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.

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 centre'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 Türkiye. With a severe physical disability. PhD in mechanical engineering from Türkiye. And some working experience there. Hobbies: adaptive sports and technologies, advocating for people with disabilities in STEM. He sits in the wheelchair, wears causal 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.

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 disable candidate -> lack of resources and obstacle for travel requirement	"Physical Disability" "Mental disability"
Age of the candidate	Negative bias towards very young candidate	

Table 14 Key takeaways from the Dutch workshop

5.3.4 1st co-creation workshop in Iceland

Intro

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:

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
Total:	**24**

Table 15 Categories and numbers of stakeholders at HI's first co-creation workshop

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 associate 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.

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"

Table 16 Key takeaways from the Icelandic workshop

5.3.5 1st co-creation workshop in Türkiye

Intro

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:

Categories	Number
HR officers/managers	3
AI specialists	3
Workers	7
Representatives of NGOs, networks, organisations fighting against discriminations	5
Total:	**18**

Table 17 Categories and numbers of stakeholders at FARPLAS' first co-creation workshop

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:

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 Türkiye, 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.

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.

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.

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 Türkiye 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 oldwith 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 Türkiye, 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 d 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 Türkiye.

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 Türkiye, as evident from the discussion regarding a candidate wearing a headscarf.

The automotive sector is usually male-dominated in Türkiye, 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.

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"

Table 18 Key takeaways from the Turkish workshop

5.3.6 1st co-creation workshop in Estonia

Intro

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:

Categories	Number
HR officers/managers	3
AI specialists	5
Workers	9
Workers' representatives	2

Representatives of NGOs, networks, organisations fighting against discriminations	1
Total:	**20**

Table 19 Categories and numbers of stakeholders at DIGI's first co-creation workshop

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:

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.

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.

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.

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.

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.

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"

Table 20 Key takeaways from the Estonian workshop

5.3.7 1st co-creation workshop in Switzerland

Intro

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:

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
Total:	**22**

Table 21 Categories and numbers of stakeholders at BFH's first co-creation workshop

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.

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 Negative bias towards Kosovar migrants -> "stigma", toxic masculinity (associated with doing weigh training) Positive bias towards Swiss candidates	"Kosovar origins"
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"

Table 22 Key takeaways from the Swiss workshop

5.4 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.

5.4.1 Overview of the collected wordlists

Overall, **389 words/sentences** have been collected by partners and included in the spreadsheets that are accessible through this link. 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.

The graph below shows the distribution of gender and race/ethnicity bias between positive and negative bias.

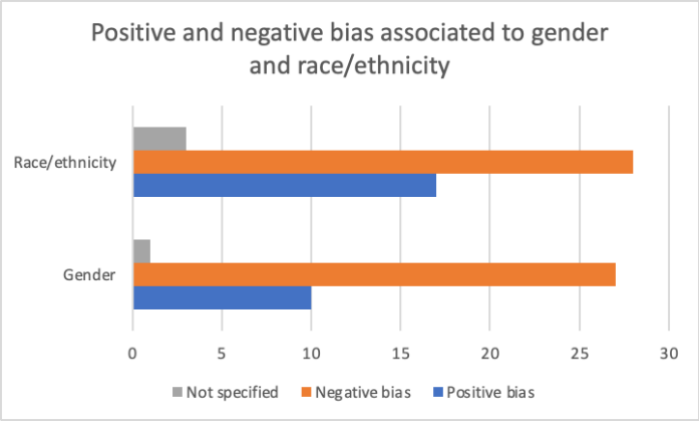

Figure 9 Positive and negative bias associated to gender and race/ethnicity

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,

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

Table 23 Other grounds of discrimination addressed by partners in the spreadsheets

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.

The graph below shows the allocation of the words/sentences among the different categories detailed above.

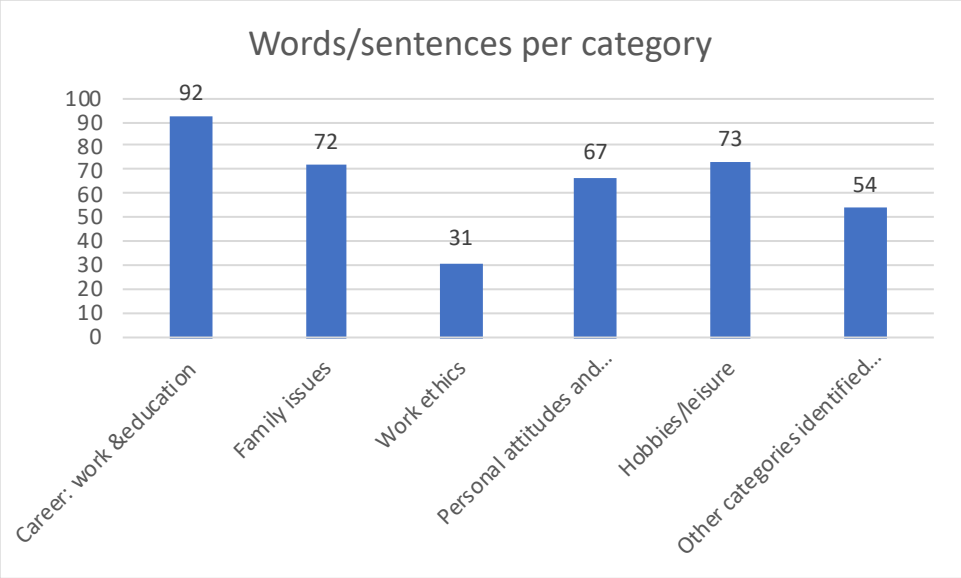

Figure 10 Words/sentences per category

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

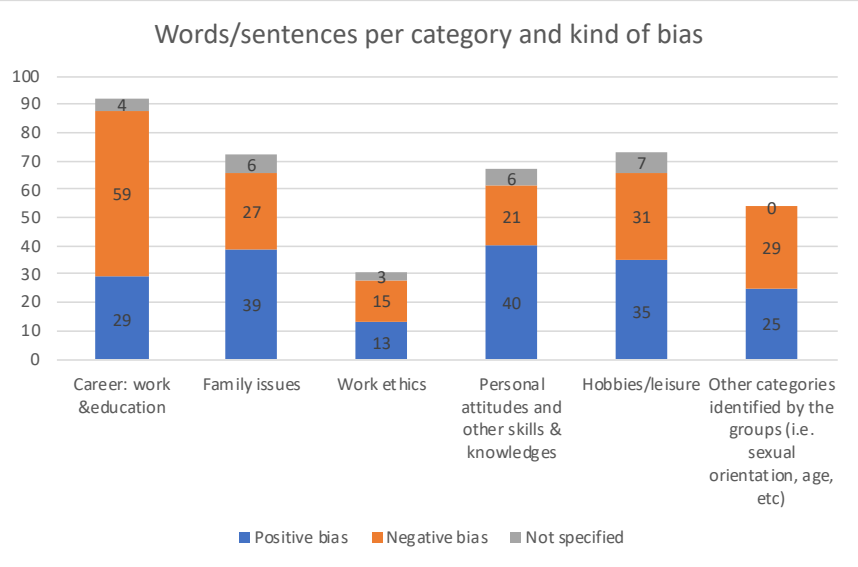

Figure 11 Words/sentences per category and kind of bias

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.

5.4.2 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.

6 Methodology of the second co-creation workshop

6.1 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 the table below:

Type	**Ideal number**
HR officers and networks, associations of HR specialists preferably already active on gender/diversity & inclusion issues	10-12
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
Legal experts in human rights and/or labour law	2
Philosophers	2
Workers and workers' representatives (e.g. trade unions)	2-4
AI specialists	4

Table 24 Categories and numbers of stakeholders involved in the second co-creation workshop

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 as follow:

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

Table 25 Structure of the second workshop

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., however this was left to the discretion of each partner to decide on.

6.2 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.

6.3 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 points[12]:

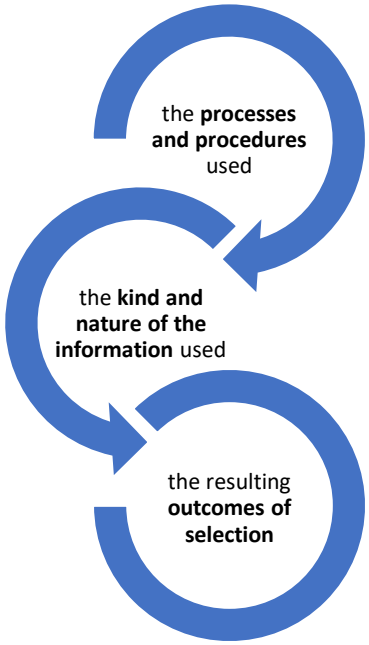

Figure 12 Main points concerning fairness in HR recruitment/selection

Regarding the first block "**the processes and procedures used**", and as pointed out in D2.1, relevant literature[13] has identified the following as the most common principles of procedural fairness:

[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),

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. Professionality -> 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.
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 etcnicity, religious belief, different abilities, etc.

1st activity: *considering principles 1, 2, 3, 5, 6, 7 (since professionality 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 this folder:

- 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 this link.

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)

6.4 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, in particular the slide sets available at this link can be used.

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.

ii. 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
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

iii. 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.

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.

Table 26 Scenario and persona of the second workshop

b. Mapping the current-state recruiter experience

[14] https://www.mindtools.com/aiwjjpy/designing-future-state-customer-journeys

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/

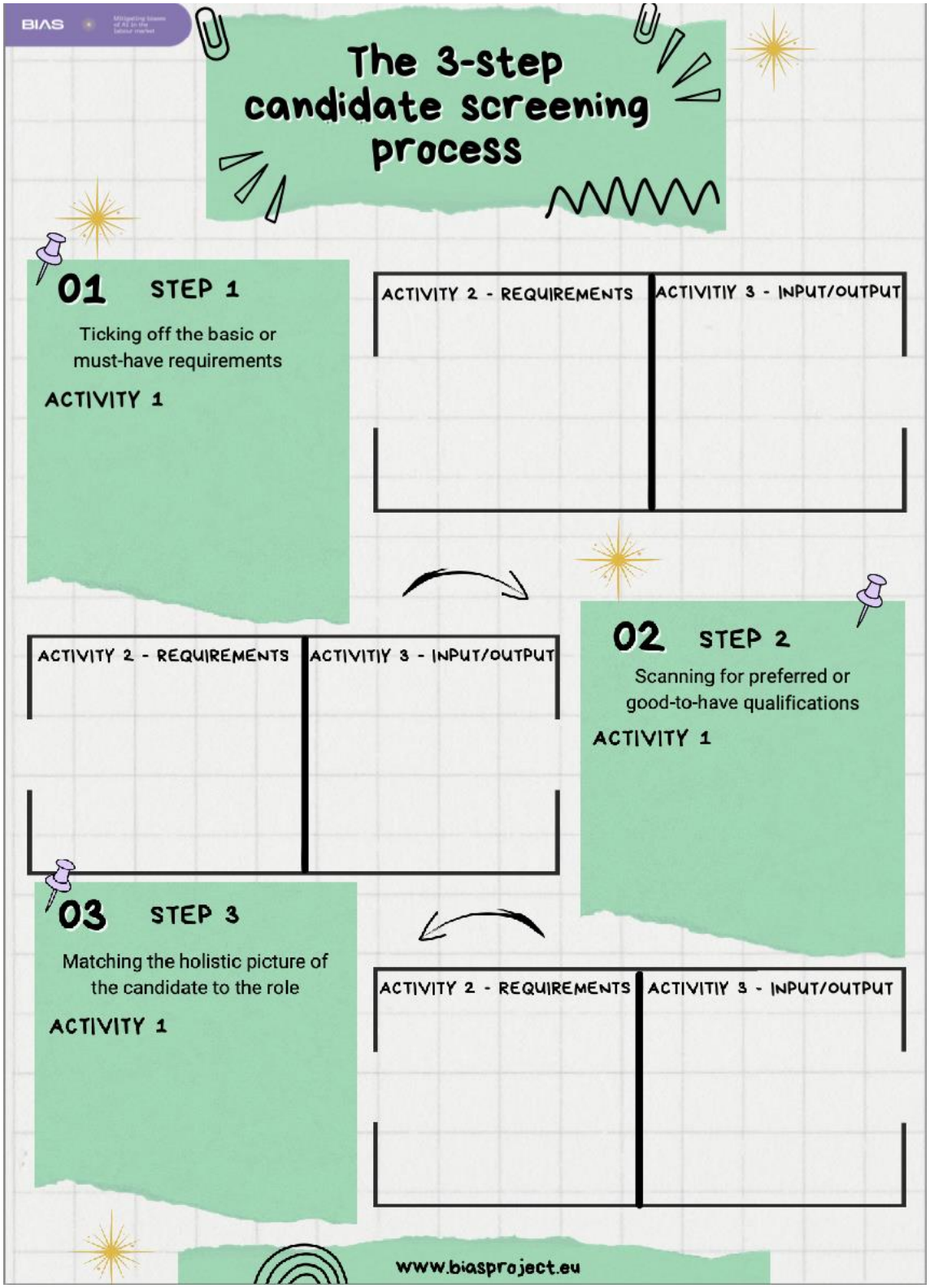

Figure 13 Poster template for the second group work exercise

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 seconds[16]. 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

[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 five steps identified above were developed as follow. After splitting the groups in four, the facilitators distributed the following materials to the groups:

- The poster with the three steps of the screening process (to be printed in A1 format) available for each partner:
  - HI
  - NTNU
  - ULEID
  - FARPLAS
  - DIGI
  - BFH
  - SVEN
- This table for reporting activity 4 (to be printed in A3 format)
- The table below with the description of the activities (to be printed in A4 format)
- The slides introducing the Debiaser and the CBR model available here
- The template for the rapporteur available here
- Sticky notes of at least 3 different colours

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

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.

Table 27 Scenario and persona of the second workshop

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.

Activity	Description
First activity – brainstorming on the screening phase (10 min)	The group observes the three steps of the screening phase and brainstorm on the following questions: - 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? 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. The rapporteur reports on the template provided making sure to also highlight any different positions/ideas of participants.

Second activity – identifying requirements (20 min)	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). In particular, per each screening step identified in the poster, the group should try to answer the following questions: - 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 indicate any sensitive information/expression in the CVs that are at risk of causing bias) - 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) 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. 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. The rapporteur fills the report indicating the different requirements identified differentiating per each screening phase as well as any diverging positions among participants.
Third activity – identifying inputs and output of the tool (10 minutes)	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. For instance, in the example provided, the needed inputs are the CVs and the job offer, while the expected outputs are: - the identified/selected CVs with the matching % between CVs' skills and job offer's skills - sensitive information/expressions and explanations 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.
Fourth activity – identifying conditions for fairness/trustworthiness and evaluation (10 minutes)	The group brainstorms on the following two points: - 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. While participants will the table provided at this **link**, rapporteurs take notes on the development of the exercise pointing out at different positions emerged among participants.
Walking plenary (30 minutes)	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.

Table 28 Structure and description of the second group work

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.

6.5 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 (the template is available here)[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

7 Current state of the second co-creation workshops

At the time the current documenwas delivered, all partners have conducted the second round of co-creation workshops, which were finalized at the beginning of October 2023. Reports from the workshops are in the making and will be analysed and presented in the second version of the present Deliverable due by January 2024.

8 Concluding remarks and next steps

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 aimed at providing input for the early development phases of the Debiaser in W3. It also summarizes and provides an initial analysis of the results from the first round of co-creation workshops conducted by seven partners in their respective countries between June and July 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 on 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 Debiaser tool and a CBR system, along with related risks. These co-creation events already took place in the various partner countries between August and early October 2023. However, reports were not available at the time this report was compiled. Therefore, the results from those reports will be included in the second version of this document, which is expected to be released by the end of January 2024. Next steps, as far as the BIAS co-creation process is concerned entail:

- Analysing results of the second round of co-creation workshops.

- Based on all these lessons learned, design of the specific methods and the techniques for a third, international co-creation workshop, aimed at further exploring and deepening desirable requirements and functionalities of a Debiaser tool and a CBR system. Date for this meeting is already set, as it will take place on the 7th of December 2023 in Venice (Italy) and will involve consortium partners and around 2 external stakeholders each.
- Drafting and submission of D2.4, due by M15 (January 2024): this report will include both the methodology for the international co-creation workshop, and results of the second round of national workshops along with an overall assessment of the integration of co-creation within WP3.

The second co-creation phase 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.

9 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 Gendel Bias in Pre-Trained Germa 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
- 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
- Saravanakumar K., The impossibility Theorem of Machine Fairness – A Casual Perspective, Columbia University, 2021
- 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.

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 dynamics[18].

The identified principles were applied to training and are selected and adapted to facilitation in co-creation settings as following:

1. self-reflection and **R**eflexivity: 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. **C**ontextualisation: 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 "**K**nowledges" 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 **T**ransformation: 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. **S**tandpoint 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. **I**ntersectionality: 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., 2022[19]).

[18] See D3.3 "Quality standards Booklet", GE Academy project, December 2021
[19] 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" [20].
Conflicts can also originate from resistances of any stakeholders to specific topics. They can be manifested in two main ways:

[20] https://www.merriam-webster.com/

- 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.[21]

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?

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)	- 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
I seek to investigate issues with participants in order to find solutions that are mutually acceptable.	**Collaborative** (conflicts are seen as problems to be solved)	- When maintaining relationships is important - When learning and trying to merge different perspectives - When time is not a concern

[21] https://www.superaproject.eu/wp-content/uploads/2022/02/Resistances-to-Structural-Change-in-Research-and-Innovation_v02.pdf

I avoid discussing my differences with	**Avoiding** (better hide and ignore conflicts and resolve it, give up personal goals and display passive behaviour)	- 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
I attempt to meet the expectation of participants	**Accomodating** (ignore own goals and resolve conflicts by giving into others)	- 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
I try to reach compromises through negotiation	**Compromising** (willing to sacrifice some goals while persuading others to give up part of theirs)	- 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

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 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 with 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.

BIAS

Mitigating biases of AI in the labour market

Consortium

NTNU

UNIVERSITY OF ICELAND

LOBA®

Crowdhelix

SMARTVENICE S.R.L.

Universiteit Leiden eLaw

Digiotouch

farplas

FH Bern University of Applied Sciences

Funded by the European Union

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

email
info@biasproject.eu

website
www.biasproject.eu

social media
@BIASProjectEU

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