BIAS

Mitigating biases
of AI in the
labour market

BIAS Co–Creation Workshops:
Insights and Results

Throughout 2023, the BIAS project conducted a **series of co-creation workshops aimed at advancing the early development of the Debiaser tool**[1] **and Case-Based Reasoning (CBR) systems**[2]**.** Organized by Smart Venice in collaboration with the Norwegian University of Science and Technology (NTNU) and Bern University of Applied Sciences (BHF), **these workshops were designed to engage a diverse range of stakeholders in refining the project's technological solutions.** The outcomes and methodologies from these workshops are detailed in D2.4, titled "Final Report on Co-Creation Methodologies and Findings."This factsheet offers a concise overview of the workshops' activities, key findings, and their impact on the progression of the BIAS project.

Workshop Details

SESSIONS:

15 **Workshops**
14 national
1 international

TIMELINE:

JUL   **2023**

SEP
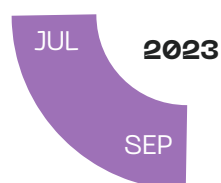
STAKEHOLDERS:

Workers,
Civil society,
HR specialists,
AI specialists,
Legal experts,
Social Sciences
and Humanities experts

PARTICIPANTS:

314 **in total**

Key Activities and Findings

National Workshops
1st Round → June–July 2023

FOCUS: Identify categories of wordlists for bias detection in word embeddings[3].

DIMENSIONS ANALYSED: Gender, ethnic/cultural background, and other potential biases.

RESULTS:

Data Collected:
389 words/sentences identified, including:
– 38 related to gender bias
– 48 related to race/ethnicity bias
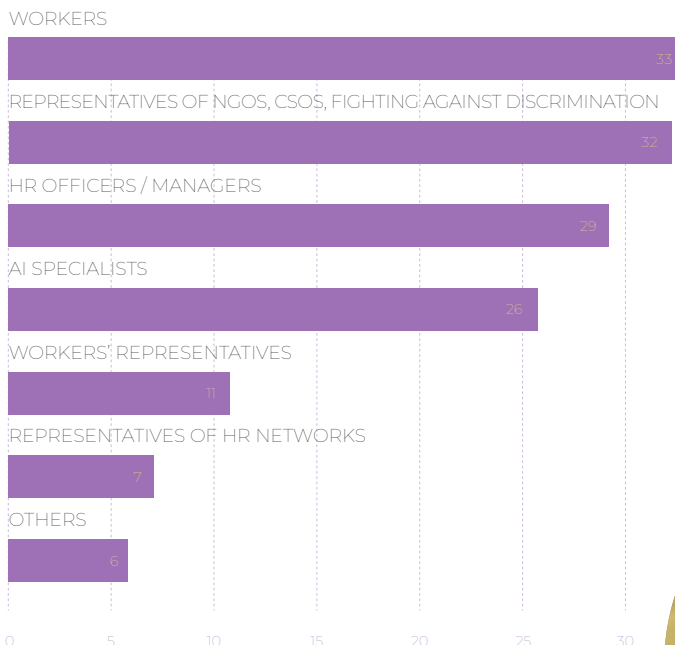– A few identified as intersectional biases

Categories:
– "Career: work & education" had the most words / sentences (92), with 59 leading to negative bias.
– "Personal attitudes and other skills & knowledge" had 67 entries, and it was the category with the higher number of words (40) leading to positive bias.
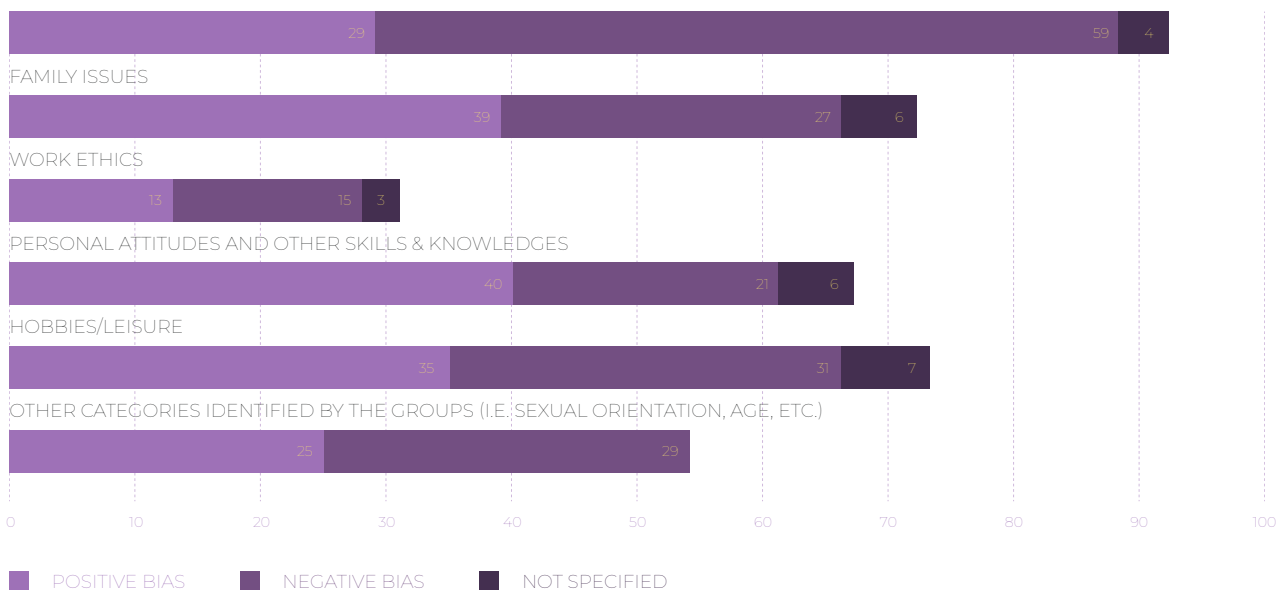
Impact:
The wordlists from this workshop will support AI experts in developing bias detection models[4].

STAKEHOLDERS' CATEGORIES:

WORKERS
33

REPRESENTATIVES OF NGOS, CSOS, FIGHTING AGAINST DISCRIMINATION
32

HR OFFICERS / MANAGERS
29

AI SPECIALISTS
26

WORKERS' REPRESENTATIVES
11

REPRESENTATIVES OF HR NETWORKS
7

OTHERS
6

0   5   10   15   20   25   30

WORDS / SENTENCES PER CATEGORY AND KIND OF BIAS:

CAREER: WORK & EDUCATION

29 | 59 | 4

FAMILY ISSUES

39 | 27 | 6

WORK ETHICS

13 | 15 | 3

PERSONAL ATTITUDES AND OTHER SKILLS & KNOWLEDGES

40 | 21 | 6

HOBBIES/LEISURE

35 | 31 | 7

OTHER CATEGORIES IDENTIFIED BY THE GROUPS (I.E. SEXUAL ORIENTATION, AGE, ETC.)

25 | 29

0 10 20 30 40 50 60 70 80 90 100

■ POSITIVE BIAS ■ NEGATIVE BIAS ■ NOT SPECIFIED

POSITIVE AND NEGATIVE BIAS ASSOCIATED TO GENDER AND RACE/ETHNICITY

RACE / ETHNICITY

GENDER

0 5 10 15 20 25 30

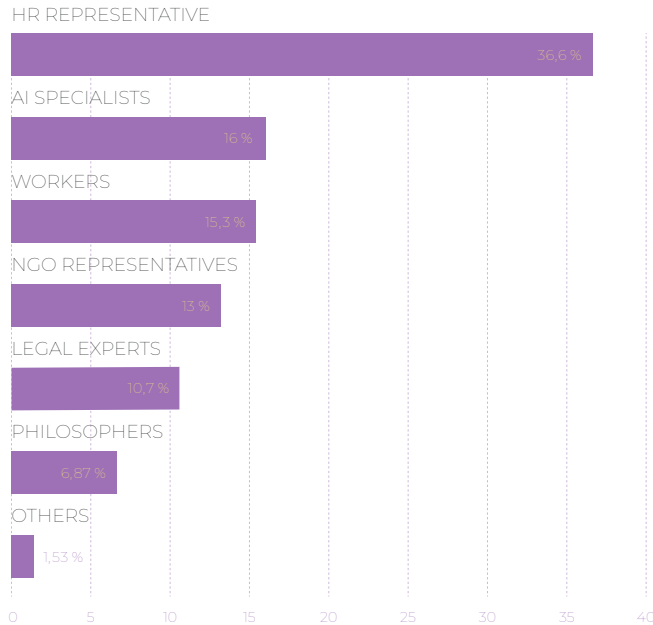■ POSITIVE BIAS ■ NEGATIVE BIAS ■ NOT SPECIFIED

NW 1
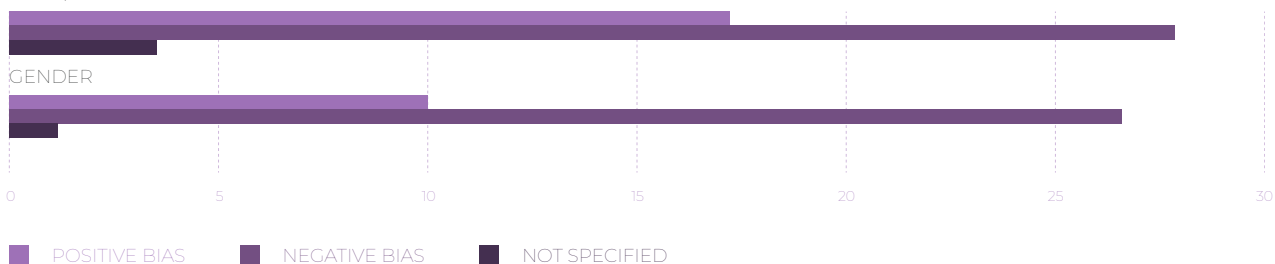
National Workshops
2nd Round → August–October 2023

FOCUS: Discuss and define fairness in HR recruitment processes.

PARTICIPANTS AND THEIR ROLES:

HR REPRESENTATIVE
36,6 %

AI SPECIALISTS
16 %

WORKERS
15,3 %

NGO REPRESENTATIVES
13 %

LEGAL EXPERTS
10,7 %

PHILOSOPHERS
6,87 %

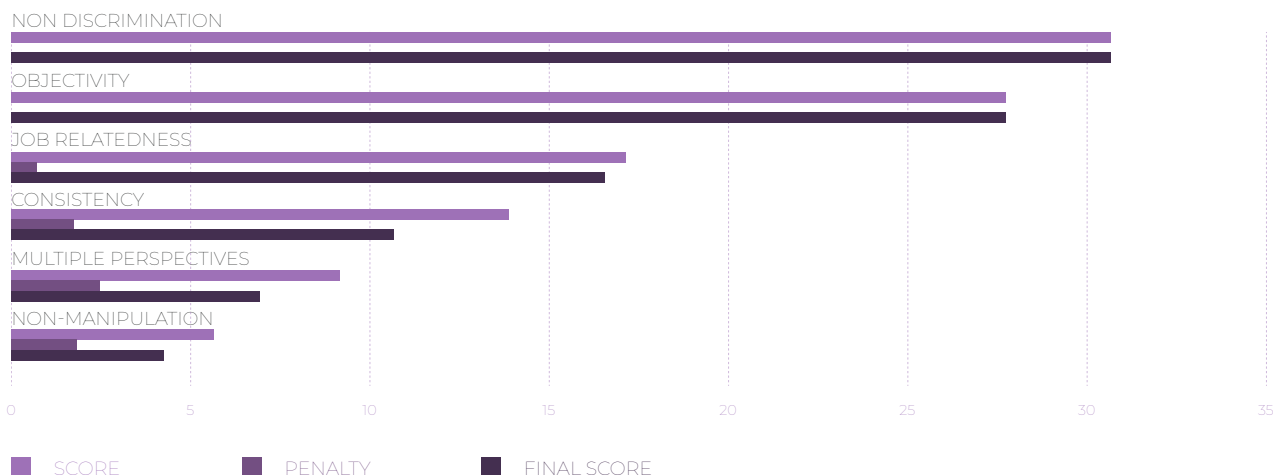OTHERS
1,53 %

0 5 10 15 20 25 30 35 40

NW 2

The second workshop provided participants with an **opportunity to discuss fairness in the initial phase of the recruitment process**. Topics included:
– Identifying fairness principles.
– Defining features of a fair recruitment process.
– Prioritising candidate attributes alongside the qualifications and skills required for a job offer.
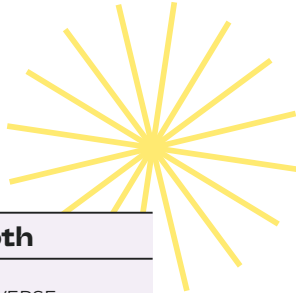
Additionally, the workshop allowed for:
– The identification of desirable requirements and functionalities for the Debiaser tool and the CBR–based decision–making support system.
– The assessment of related risks.
The discussion centred on what a fair HR recruitment process should entail, resulting in a set of principles that were discussed, rated, and subsequently presented in the "**Ranking of Principles Discussed**" graphic.

NON DISCRIMINATION

OBJECTIVITY

JOB RELATEDNESS

CONSISTENCY

MULTIPLE PERSPECTIVES

NON-MANIPULATION

0 5 10 15 20 25 30 35

■ SCORE ■ PENALTY ■ FINAL SCORE

NEEDS AND REQUIREMENTS FOR AI TOOLS — NLP BASED AND CBR DEBIASER TOOLS:

During the second co-creation workshop, participants focused on defining the essential requirements and functionalities for two systems: the Debiaser tool and the CBR-based decision-making support system. The discussion also covered potential risks associated with each system. The table below illustrates the outcomes, organised by different sections such as Business Needs, Screening Phase, and others. Each section identifies specific needs and functionalities that are unique to the CBR system, the Natural Language Processing (NLP) system[5], or relevant to both systems. This helps clarify which features are best suited for each system and where they overlap.

Section	CBR	NLP	Both
BUSINESS NEEDS	- FILL VACANT POSITIONS - OPTIMAL MATCHING - SIMPLIFY HIRING PROCESS	- REDUCE APPLICATION SCREENING TIME	- COMPARE DIVERSE APPLICATIONS EASILY
SCREENING PHASE	- DIFFERENT OUTPUTS PER PHASE - INTEGRATE PHASE OUTCOMES - RECORD FINAL DECISIONS	- AI SUPPORTS, DOESN'T DECIDE - RESTRICT AI TO INITIAL PHASE	- COMPLIANCE WITH POLICIES AND REGULATIONS
SYSTEM INPUT	- PRE-STRUCTURED CVS - BLIND CV FORMATS - COVER LETTERS AND RESUMES	- INCLUDE SOCIAL PROFILES	- SUPPLEMENTARY "KILLER QUESTIONS" INPUTS
SYSTEM OUTPUT	- GENERATE INTERVIEW SHORTLISTS - IDENTIFY COMPENSATION BENCHMARKS - ALTERNATIVE REQUIREMENTS	- GENERATE INTERVIEW SHORTLISTS - INDICATE PREFERENCES	
SYSTEM USAGE	- TRAINING PROGRAMS - ADAPT TO REGIONAL REGULATIONS	- INTUITIVE INTERFACE - EFFICIENT USAGE	
MODEL TRAINING	- DIVERSE DEMOGRAPHIC SAMPLES - NON-DISCRIMINATORY DATA - CROSS-VALIDATION	- ADAPTIVE LEARNING	
ENSURING FAIRNESS	- FOCUS ON ESSENTIAL SKILLS		- FOCUS ON ESSENTIAL SKILLS - COMPLY WITH DATA PROTECTION - AVOID IRRELEVANT BIASES - PERIODIC REVIEWS - TRANSPARENCY AND USER AWARENESS

CONCLUSIONS: Reflecting on the outcomes of the exercise and the feedback provided by the groups at the conclusion of the workshop cycle, it became evident that the two tools, the Case-Based Reasoning (CBR) system and the NLP system, were perceived with distinct strengths. The CBR tool was deemed more suitable for the first phase of screening, while the NLP-based system emerged as a versatile, cross-phase tool, with particular efficacy during a later phase of the recruitment process. This is the phase where recruiters aim to achieve a holistic match between a candidate and the company, as it is at this stage that recruiters are more susceptible to potential biases, having to carefully analyse all the textual information about candidates.

Also, from the second round of workshops, it became evident that a more practical approach was necessary for stakeholders to give their relevant feedback on the needs and requirements for the Debiaser.

International Workshop
7th December 2023

PARTICIPANT DISTRIBUTION:

PARTNER ORGANIZATION REPRESENTATIVES
52,5 %

EXTERNAL STAKEHOLDERS
47,5 %

0   10   20   30   40   50   60   70   80   90   100

FOCUS: Simulate AI-based recruitment systems, discuss trustworthiness, and gather learning requirements.

STRUCTURE:
– **Simulating AI Systems:** Participants used the **Candidate Ranker** and **Mitigation Tool** to evaluate AI's role in hiring.
– **Trustworthiness:** Discussions aligned system requirements with the ALTAI[6] ethical guidelines.
– **Learning Needs:** The BIAS consortium collected ideas for developing effective training packages for AI tools.

The participants were previously provided with a real-life scenario where they had received a job advertisement, a list of features, and some questions via email, and had the opportunity to reflect on them. Participants were later asked to configure the tools and evaluate their functionalities. Both simulation tools incorporated ChatGPT as the reasoning component, aligning with the primary goal of encouraging participants to engage actively with AI tools.

These exercises aimed to gather feedback on the tools and establish specific requirements and details.
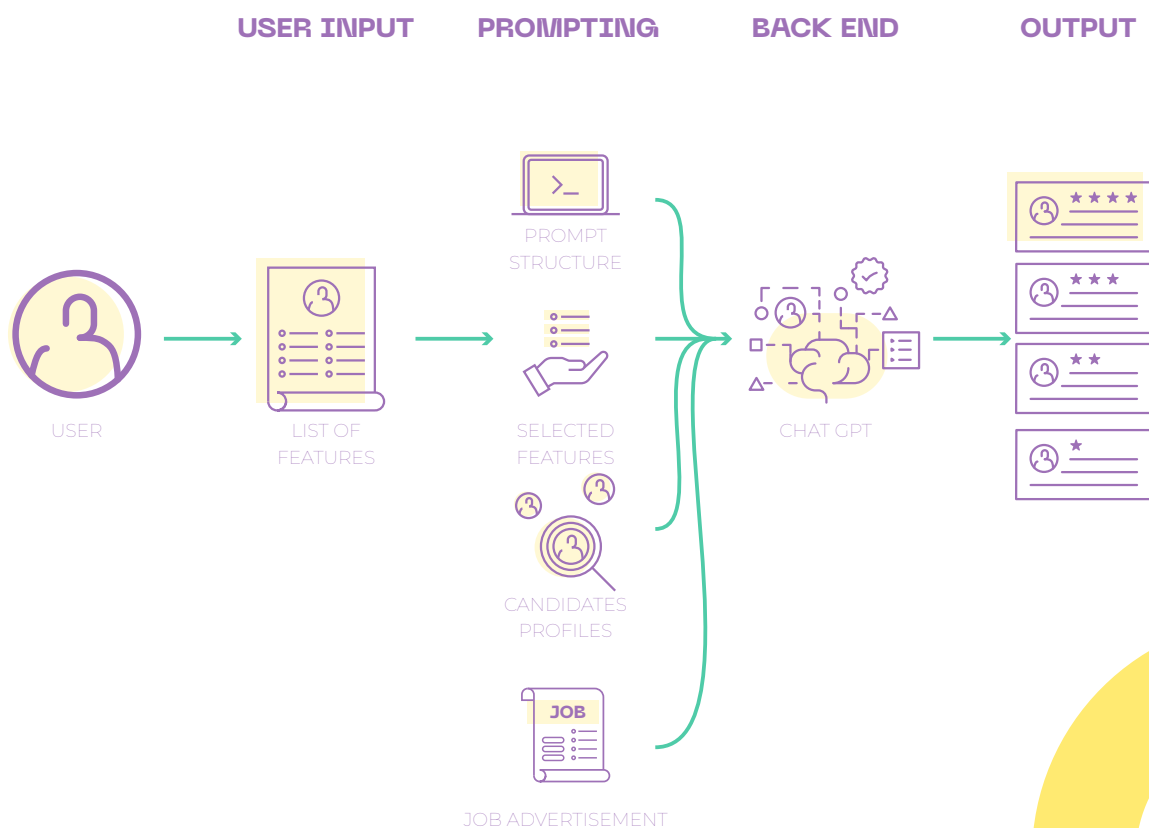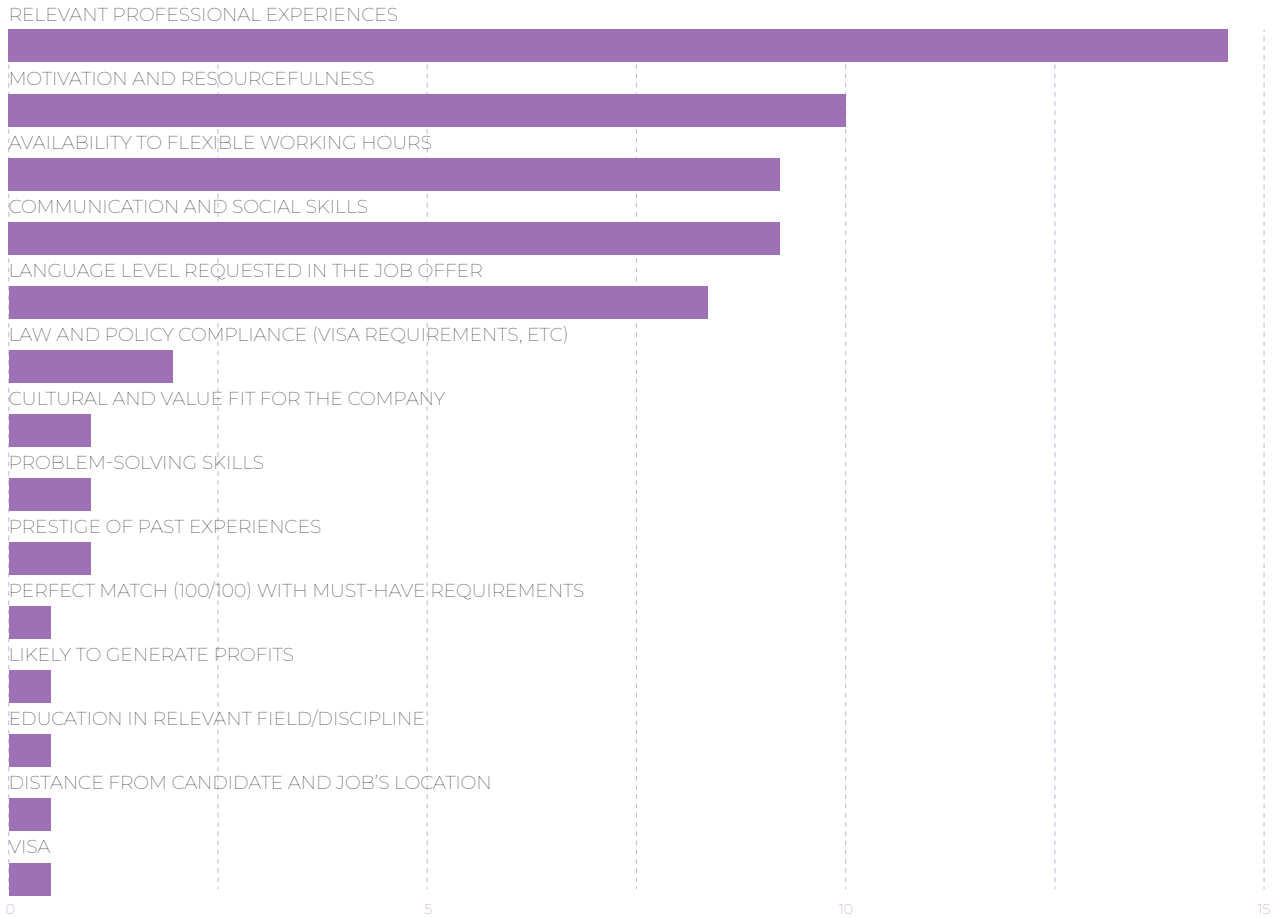
FINDINGS:

Candidate Ranker:

→ A simulator that emulates the CBR/NLP system for selecting suitable candidates.

  – **Strengths:** Successfully prioritise features, though transparency in feature influence was needed.

  – **Issues:** Potential biases, lack of human oversight, and need for data anonymization.

  – **Improvements:** Suggested balancing AI support with human decision-making for fairness and accuracy.

THE "CANDIDATE RANKER":

| USER INPUT | PROMPTING | BACK END | OUTPUT |

MUST-HAVE REQUIREMENTS SELECTED:

RELEVANT PROFESSIONAL EXPERIENCES

MOTIVATION AND RESOURCEFULNESS

AVAILABILITY TO FLEXIBLE WORKING HOURS

COMMUNICATION AND SOCIAL SKILLS

LANGUAGE LEVEL REQUESTED IN THE JOB OFFER

LAW AND POLICY COMPLIANCE (VISA REQUIREMENTS, ETC)

CULTURAL AND VALUE FIT FOR THE COMPANY

PROBLEM-SOLVING SKILLS

PRESTIGE OF PAST EXPERIENCES

PERFECT MATCH (100/100) WITH MUST-HAVE REQUIREMENTS

LIKELY TO GENERATE PROFITS

EDUCATION IN RELEVANT FIELD/DISCIPLINE

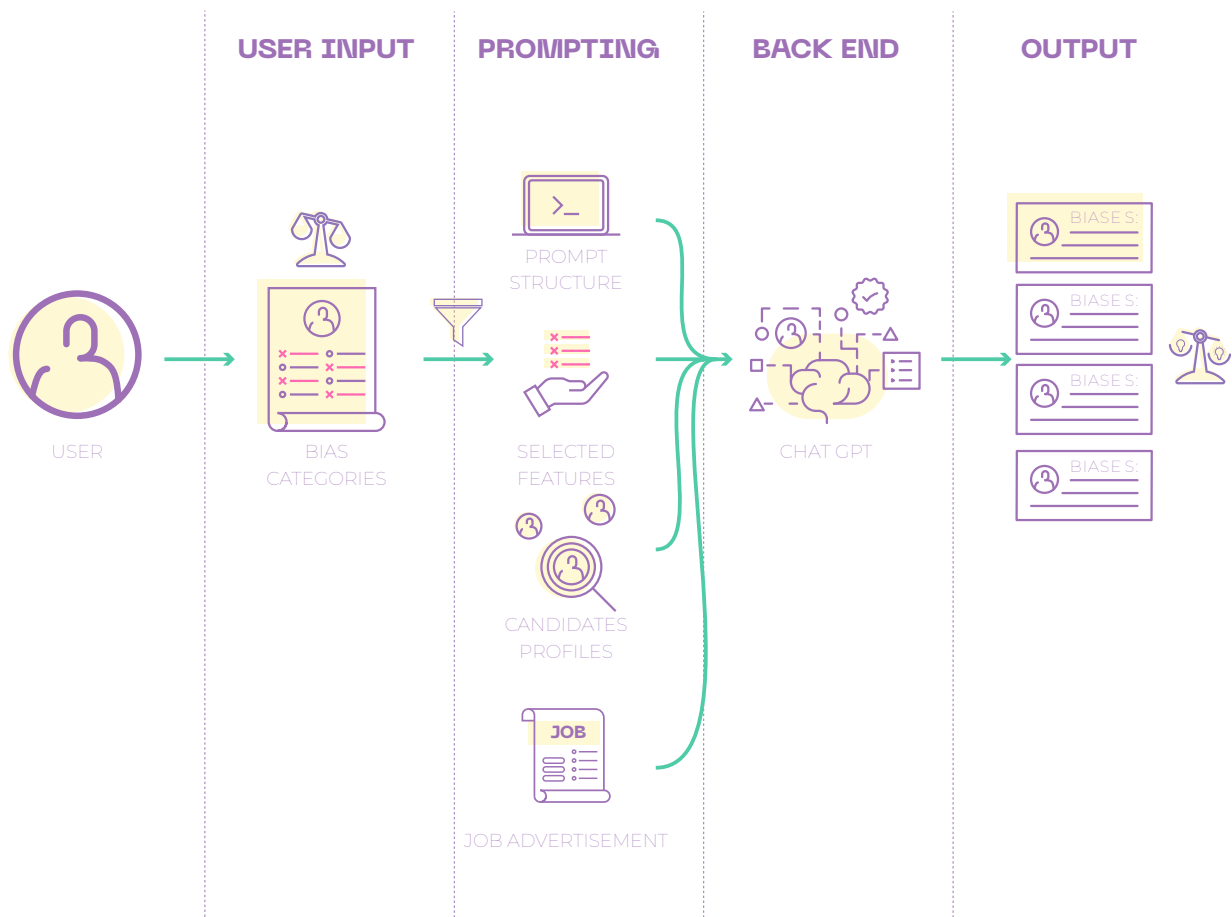DISTANCE FROM CANDIDATE AND JOB'S LOCATION

VISA

After prioritising features, participants observed the Candidate Ranker's outcomes. The tool easily identified the "best" candidate but struggled with comparing the relative merits of all candidates, as highlighted by the chart below.

GAUSSIAN DISTRIBUTION[7] OF RANKING POSITION OF CANDIDATES BY THE SIMULATING TOOL
(MEAN AND VARIANCE)

CANDIDATES:

- ANNA
- MOHAMMED
- MARK
- DIEGO
- PRIYA
- FELIX

FREQUENCY
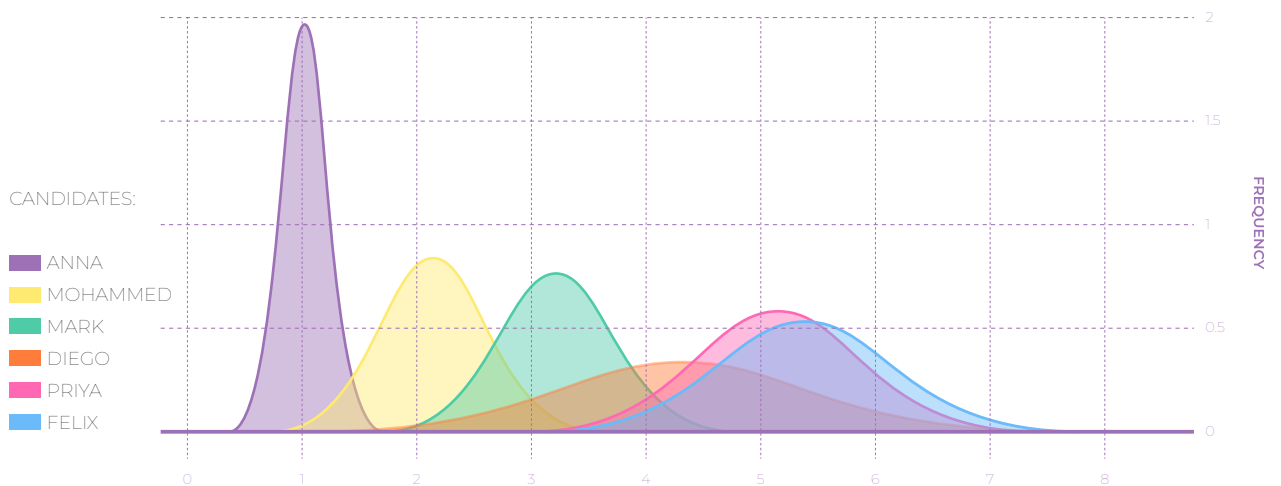
MITIGATION TOOL:
→ The Mitigation Tool aimed to simulate the word–embedding–based system that identifies potential biases during profile screening.
– **Strengths:** Effective in identifying biases, but had inconsistencies.
– **Issues:** Tool assumptions and potential reinforcement of biases.
– **Recommendations:** Avoid using photos, improve explanation clarity, involve social linguists and psychologists.

| USER INPUT | PROMPTING | BACK END | OUTPUT |

USER

BIAS CATEGORIES

PROMPT STRUCTURE

SELECTED FEATURES

CANDIDATES PROFILES

JOB

JOB ADVERTISEMENT

CHAT GPT

BIASE S:
BIASE S:
BIASE S:
BIASE S:

TRUSTWORTHINESS:
→ This theoretical session was strategically scheduled after participants had interacted with a simulation of the systems, providing them with a clearer understanding to analyse the requirements more effectively.
– Key issues included fairness vs. trustworthiness, system limitations, privacy, security, human oversight, and accessibility.
– Need for robust protections, simpler algorithms, and better user training emphasised.

DEVELOPMENT IMPACT

The insights gained from these co-creation workshops were pivotal in shaping the proof-of-concept technology for the BIAS project. By involving stakeholders from diverse backgrounds, the project ensured a user-centred development process that addressed the needs of various potential users, thereby enhancing the effectiveness and trustworthiness of the Debiaser tool.

1 – DEBIASER TOOL

The BIAS project's innovative technology designed to identify and reduce biases in language models used in AI. These models, trained on large text datasets, can sometimes reflect societal biases. The Debiaser tool helps detect and adjust these biases, making the models safer and fairer for use in areas like Human Resources Management, where unbiased decisions are crucial.

2 – CASE-BASED REASONING (CBR) SYSTEMS

AI technique that solves new problems by reusing solutions from similar past problems. A CBR system has two main components: the Case Base, which stores previously solved problems, and the CBR engine, which retrieves similar cases and applies their solutions to new problems. This process helps ensure "individual fairness" by evaluating similar individuals in a consistent way.

3 – WORD EMBEDDINGS

A type of word representation in AI where each word is assigned a vector (numerical representation). This allows easier processing of words and language for computation.

4 – BIAS DETECTION METHODS

Methods designed to identify and measure biases in data or models. They help detect unfair patterns or discrimination, ensuring more balanced and equitable outcomes in decision-making and other AI systems.

5 – NATURAL LANGUAGE PROCESSING (NLP) SYSTEM

Systems that enable computers to understand, interpret, and generate human language.

6 – ALTAI

The Assessment List for Trustworthy Artificial Intelligence (ALTAI) is a tool developed by the European Commission to help organizations assess and improve the trustworthiness of their AI systems. It is part of the European Union's efforts to ensure that AI is developed and deployed in a way that is ethical, legal, and robust. ALTAI provides a structured framework for AI self-assessment across several key areas, encouraging developers to evaluate their AI systems based on principles of 1) human agency and oversight, 2) technical robustness and safety, 3) privacy and data governance, 4) transparency, 5) diversity, non-discrimination and fairness, 6) environmental and social well-being, 7) and accountability.

7 – GAUSSIAN DISTRIBUTION

A Gaussian distribution chart, also known as a normal distribution chart, visually represents a bell-shaped curve that is symmetric around its mean. It is widely used in statistics and probability theory to model data that clusters around a central value with no bias to the left or right.
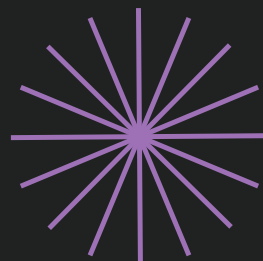
Visit us

www.biasproject.eu

Contact us

info@biasproject.eu

Follow the BIAS project

f in ▶ 🐦 �📷 ◢   @BIASProjectEU

BIAS Project Partner Organisations

NTNU

UNIVERSITY OF ICELAND

LOBA®

Crowdhelix

SMARTVENICE srl

Universiteit Leiden eLaw

Digiotouch

farplas

FH Bern University of Applied Sciences