# Fairness trade-offs in hiring: what people prefer and what engineers can build

Pavlo Burda
ICT Institute
Utrecht, Netherlands
pavlo@ictinstitute.nl

Sieuwert van Otterloo
HU University of Applied Sciences
Utrecht, Netherlands
sieuwert.vanotterloo@hu.nl

## Abstract

Human-centered AI must confront tensions between mutually incompatible fairness definitions and fairness requirements of algorithmic decision-making (ADM) systems. To investigate how people perceive this trade-off and how this perception can guide engineering requirements, we determine the underlying principles of common fairness metrics in the form of statements that people may or may not agree with. Using an illustrative dataset, we show how favored metrics can conflict in practice, underscoring the need for explicit trade-offs and how to solve them. We design and evaluate a survey that can be used to determine the preferences of stakeholders in a hiring scenario by mapping 12 statements to demographic parity, equal opportunity (TPR), predictive equality (FPR), predictive parity (PPV), fairness through unawareness, and individual fairness definitions. Responses (N=51) indicate broad support for excluding sensitive attributes and for error-rate parity criteria (FPR-TPR), with contrasting views on demographic parity under unequal base rates. We contribute a requirements-elicitation approach that can be used to define 'fairness requirements' of an ADM system by mapping stakeholder preferences to concrete metrics, yielding a pragmatic set of recommended requirements using our hiring scenario as a guiding example.

## Keywords

fairness, decision-making, software requirements

**ACM Reference Format:**
Pavlo Burda and Sieuwert van Otterloo. 2026. Fairness trade-offs in hiring: what people prefer and what engineers can build. In *Human Centred Artificial Intelligence - Education and Practice (HCAI-EP '26), January 21–22, 2026, Kildare, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3777490.3777496

## 1 Introduction

As machine learning algorithms are increasingly utilized to make or support significant decisions across various domains, such as recruitment, loan approval and fraud detection [10, 19], researchers and practitioners highlighted issues of bias and a lack of fairness in algorithmic decision-making [2, 3, 7]. Research proposed a multitude of fairness definitions [21], but there is no definitive consensus on which fairness definitions might apply best to obtain a 'fair decision' in given situations [18]. Moreover, it has been shown that many fairness definitions are incompatible between them [3, 11] and can be contradictory [1], thus requiring difficult trade-offs [9].

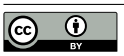
One of the affected domains by the 'fairness dilemma' is recruitment. Companies might want to implement algorithmic decision-making (ADM) systems to efficiently screen job applications to identify top candidates while mitigating biases in the process. Together with regulatory needs for high-risk AI systems, such as the EU AI Act [14], companies encounter significant challenges in defining what 'fair' means for their system: how do they translate these complex trade-offs into a system that is transparent, accountable, and intuitively perceived as fair by a diverse workforce and society?

From the point of view of software engineering, the requirements for a fair ADM can be thought of as software requirements, whereby developers elicit the requirements from users, customers and stakeholders [20]. Value Sensitive Design is a popular approach to technological design based on the idea of accounting for human values in the design process [6]. This method can be used to make sure that fairness and absence of bias are accounted for. This design method often includes an empirical investigation, where designers try to collect information from actual people on what values must be incorporated in a (ADM) system.

In our view, which is based on value sensitive design priciples, it is thus necessary for any concrete decision algorithm to find out what the exact fairness requirements are, in order to select suitable metrics. There are many possible ways to elicit requirements, and a popular way to do this is via surveys or structured interviews [20]. In this paper, we present a list of possible requirements to include in a survey or to discuss with stakeholders, and a mapping that shows which metrics to use depending on which requirements the stakeholders agree with. Our survey measures user agreement with statements that represent the principles underlying a subset of common fairness metrics in a fictitious hiring scenario.

We develop and evaluate the survey based on a plausible hiring scenario with contextualized fairness definitions in plain-language statements. Based on the outcomes with 51 participants, we report the top four statements that we recommend to ask for fixing a practical set of requirements for the example algorithmic hiring system. Our data shows that these are the requirements that many respondents think are important for fairness in a hiring scenario. The full survey is included as supplementary material and can be used for practical requirement-elicitation with users and stakeholders.

The paper is structured as follows: in Section 2 we discuss relevant work, in Section 3 we outline our approach, in Section 4 we showcase the survey outcomes and Section 5 provides our recommendations and discusses our findings.




*HCAI-EP '26, Kildare, Ireland*

ACM ISBN 979-8-4007-2153-3/26/01
https://doi.org/10.1145/3777490.3777496

## 2 Background

### 2.1 Fairness metrics

Research has proposed an abundance of definitions of fairness in ADM. We rely on the selection from previous work that collected and discussed the various known fairness metrics [1, 2, 21].

We can roughly divide the metrics into individual and group fairness [1, 2]. Individual fairness definitions share the principle that algorithms should predict similar outcomes for similar individuals. This can be implemented by means of similarity or distance metrics which, however, are difficult to apply in practice due to the lack of a meaningful, unbiased way to define a similarity metric for, e.g., ethnicity, gender and or other sensitive attributes [1, 21]. Given our practical context, we include in our study the notion of individual fairness in terms of objective attributes, such as performance scores, and fairness through unawareness of sensitive attributes [2, 12].

Group metrics are based on the idea that groups defined by a sensitive attribute (e.g., age, gender or ethnicity) might be discriminated, and fairness metrics should be evaluated at such group level: metrics are compared between a protected group (sharing sensitive attributes) and an unprotected group (the rest). Out of more than 20 definitions evaluated by Verma and Rubin [21], notable group fairness definitions include Demographic parity, Predictive parity (Positive Predicted Value or precision, PPV), Predictive equality (False Positive Rate balance between groups, FPR), Equal opportunity (True Positive Rate balance, TPR), Equalized odds (equal FPR and TPR) and Treatment equality (equal True Negative/FP ratios).

Other types of fairness definitions include metrics based on predictive probabilities, like test fairness, and causal reasoning, such as fair inference [2, 21]. These are not always practical, e.g., test fairness requires prediction probabilities at various frequencies instead of just predicted outcomes, which adds further complexity, while expert opinion is often required for measuring fair inference.

In this work, we focus on Demographic parity, Predictive parity (PPV), Predictive equality (FPR) and Equal opportunity (TPR) due their popularity and easy to follow demonstrations as highlighted in previous work [1, 21, 22].

### 2.2 Metrics in practice

To illustrate an example of metrics usage and highlight the practical conflicts between the metrics, we compute the metrics on a dummy dataset where 225 job candidates apply to a fictitious position as a security guard [1]. The applicants' records contain age, gender, address, test score (physical speed and strength), a truly qualified label, if hired by expert (perfect classifier, ground truth) and classification outcomes of four ADM systems (selected/not selected).

Let gender be the sensitive attribute, female candidates the protected group and male candidates the unprotected group. For Demographic parity to be satisfied, individuals in both protected and unprotected groups should have similar probability of being selected by the classifier. For Predictive parity, the algorithm precision should be similar for both groups. For Predictive equality (FPR), the probability of an unqualified applicant to be selected by the algorithm is similar between the groups. By similar quantities we mean the difference between the two quantities is below a certain threshold, in our case 0.1 or 10%.

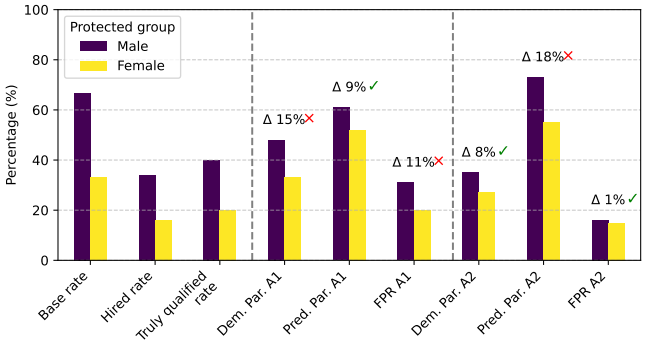



**Figure 1: Base rates and fairness metrics applied to the example dataset [1].**

**Table 1: Compatibility matrix between the selected metrics. ○- typically not compatible, ◐- compatible under rare conditions, ●- typically compatible.**

| | Dem. par. | Unawar. fair. | Eq. opp. (TPR) | Pred. eq. (FPR) | Pred. par. | Ind. fair. |
|---|---|---|---|---|---|---|
| **Dem. par.** | - | | | | | |
| **Unawar. fair.** | ○ [4] | - | | | | |
| **Eq. opp. (TPR)** | ○ [3] | ○ [4, 8] | - | | | |
| **Pred. eq. (FPR)** | ○ [3] | ○ [4, 8] | ● [2, 21] | - | | |
| **Pred. par.** | ○ [3, 11] | ◐ [4, 15] | ○ [3, 8, 11] | ○ [3, 8, 11] | - | |
| **Ind. fair.** | ○ [4] | ● [4] | ◐ [4, 15] | ◐ [4, 15] | ◐ [3, 15] | - |

Figure 1 illustrates the application of Demographic parity, Predictive parity and Predictive equality (FPR) definitions of hiring outcomes from the example dataset on two different algorithms A1 and A2. The base rate of male and female applicants is 66% and 33% respectively, the rates of hired candidates is 34% and 16%, and the rates of truly qualified candidates (necessary to compute Predictive parity and Predictive equality, FPR) are 40% and 20%. In the middle of Figure 1, we show the three metrics according to A1 algorithm that select the candidates for hiring in the dataset: Demographic parity and Predictive equality (FPR) are not satisfied as the differences ($\Delta$) between the quantities for protected and protected groups are greater than 10% (as defined in [1]); only Predictive parity with a $\Delta \leq 10\%$ is. A2, on the other hand, would be fair according to Demographic parity and Predictive equality (FPR), but not Predictive parity. Other metrics such as individual fairness and fair inference are not practically computable on the dummy dataset: there is no natural way to build an individualized profile of each candidate for the former, and there is no causal explanation of the A1 and A2 decisions for the latter.

Table 1 helps us to visualize a 'compatibility matrix' of all selected metrics. Most of them are incompatible between each other, with some compatible under rare conditions (equal base rates, near-perfect predictions) such as Predictive parity and Fairness through unawareness when a sensitive attribute is mostly not recoverable from other features [4, 15], or Individual fairness with FPR-TPR depending on the used similarity metrics [4, 8, 15].

## 2.3 Fairness as software requirement

The example above points to the fact that implementing even simple metrics leads to non-trivial trade-offs on which metric to implement in a given ADM system. Let us consider a company that desires to implement an ADM system to help with screening and hiring job applicants. Among the various requirements, such as reliability and usability, the company strives for the system to be fair towards candidates. Here arises the practical need to precisely define a set of concrete software requirements for the software engineers. This is at odds with the example above where it can be very difficult for the stakeholders to decide which fairness definitions should be used as technical specifications for the hiring system.

There is, indeed, a growing need of practical and fair implementations in business, e-governance and general software design, necessitating a careful requirement elicitation [20]. Implementing ADM systems with concerns for fairness, can draw upon approaches like Values Sensitive Design (VSD) which integrates human values into the technological design process, beyond traditional criteria like usability and reliability [6]. In VSD, a conceptual investigation of values important to the stakeholders (i.e., applicants and business) is integrated into the technical specifications of the ADM. This can be, for instance, a set of user interface indicators to show how an AI system takes a given decision [13]. A practical way to gather values central to the stakeholders, including fairness, is by observing or measuring the human context in which the ADM system operates by means of surveys or interviews. This approach substantially overlaps with requirement elicitation methods often used in software engineering: software requirements are a detailed descriptions of the functions, features, and constraints that a software system must possess to meet a user's needs or solve a specific problem. Surveys are meant to collect user needs or perceptions with respect to a desired functionality and therefore are a valuable method to elicit requirements from a large number of people [20].

## 2.4 Related work

Several empirical studies investigated human perceptions of fairness in ADM [5, 7, 9, 17–19] with some focusing on hiring-related decisions [10, 17]. For example, a study investigating perceptions of fairness in a recruitment system found that participants pointed at sensitive attributes like gender and age as the most likely cause of unfairness [10]. While favoring the use of more pertinent factors (e.g., a test score), the same participants acknowledge that this is often not sufficient for an ADM to be perceived as fair [10]. This and other studies [1, 17] point out that human perception of fairness is strongly context-dependent with substantially little chances for ironing put all possible disagreements.

These investigations suggest that there is still a open and lively discussion at crossroads of fairness metrics and practical needs of 'fair' ADMs. In this work, we approach this gap by constructing and evaluating a survey to elicit fairness requirements in an example scenario, by providing an example set of recommended requirements in our hiring scenario based on the analysis of survey outcomes, and an informal infographic of fairness statements to help guide fairness requirements discussions in practice.

# 3 Methodology

## 3.1 Survey design

We first identified, based on the literature reviewed in previous research [1, 2, 21], potential fairness metrics. We then formulated statements that capture the underlying principle of each fairness metric. The statements are an implementation in simple terms of the selected fairness metrics contextualized within an example scenario. These statements should be seen as potential requirements that can be used when implementing a recruitment procedure. We used the statements in a survey to validate that each statement is indeed understandable and usable in a survey and recognized by a broader audience as a relevant requirement for this domain. Surveys are an efficient technique to gather requirements from users and stakeholders, especially from a large set of people [20]. At the same time, measuring people preferences allows to integrate the human context - central to fairness - in which the algorithmic hiring process operates [6].

To contextualize participants' judgments of fairness statements, the survey is based on a plausible hiring scenario whereby candidates applying for a night security role are evaluated on physical tests and other attributes. The fictitious company is interested in fine-tuning the hiring process and selection algorithm to make sure the process is effective (suitable candidates are hired) and fair.

First, participants are asked 6 closed background questions comprising demographics and previous experience. Next, in Table 2 users are asked to choose whether they agree or disagree with 12 statements covering the selected fairness metrics on a 5-point Likert scale from "Strongly disagree" to "Strongly agree". An example statement is "*S3) Whether candidates are hired should not depend on age and gender*" which maps to the definition of Fairness through unawareness [21]. Each statement is an application of a metric definition to the hiring scenario in the simplest form possible. Finally, the last open question asks participants' opinion on the survey and how companies can make their hiring processes more fair. See the online supplementary material for the full scenario and survey [1].

Following best practices [16] and to ensure construct validity, the statements were derived directly from formal mathematical definitions of fairness metrics, and refined through three iterations between the investigators with expertise in algorithmic fairness and HCI. The first iteration included at least two version for each metric and a numerical example. After the investigators reached agreement, the shortest and simplest statements were chosen, with no numerical example, to not overload the participants [19]. We conducted cognitive pretesting with an external collaborator to gather feedback on statements' correct interpretation resulting in only minor adjustments.

## 3.2 Experiment design

We recruited participants over several channels focused on the Netherlands and EU in general: a class of first-year master students with non-STEM background at a Dutch university (the non-mandatory and non-graded survey was posted on the educational CMS on the intro day of the course), a newsletter targeted at more

[1] Supp. material - https://github.com/paolokoelio/hcai2026_supplementary_material

**Table 2: Fairness definitions mapped on the survey statements (S1 to S12 define the order in the published survey).**

| Definition | Statement ("Do you agree with...") |
|---|---|
| Diversity goal | S1) If multiple positions are available, the company should aim at hiring both male and female candidates. |
| Demographic parity[2, 4] | S2) The probability of a candidate being hired should be the same among male and female candidates.<br>S6) If more men than women apply to the position, it is fair that more men than women are hired. |
| Fairness through unawareness [12]<br>age & gender<br>residence<br>test-score | S3) Whether candidates are hired should not depend on age and gender.<br>S4) Whether candidates are hired should not depend on their age.<br>S5) Whether candidates are hired should not depend on their address where they live.<br>S9) The selection algorithm used by the night guard security company should only consider a candidate's test score and no other characteristics such as background, experience, age, gender or address. |
| Equal opportunity (TPR) [3] | S7) Among all individuals who are genuinely qualified based on a test result, the probability of being selected by the algorithm should be equal. |
| Predictive equality (FPR) [3] | S8) Among all individuals who are genuinely unqualified, the probability of being incorrectly selected by the algorithm should be equal for both men and women. |
| Predictive parity [3] | S10) Among all individuals that the algorithm selects, the probability that these individuals are actually qualified should be equal for female and male candidates. |
| Individual fairness [4] (≡ and ≈ similarity) | S11) If two candidates have exactly the same experience and test score, they should have the same chance of being hired.<br>S12) If two candidates have similar experience and test score, the probability of being hired should be similar. |

than 500 professionals in IT (business owners, IT consultants, workers with IT background), on social media (LinkedIn and local groups on Reddit) and personal network.

In line with ethical best practices [16], the survey was anonymous and no monetary incentives were present; the survey introduction explained the purpose of the study, who are the researchers, the voluntary nature of participation, that no tracking, no personal data collection, and no negative consequences for not filling the survey. The research was approved by the privacy officer and a contact email was provided for any questions. Our opportunistic sample amounts to 51 participants.

### 3.3 Data analysis

From the gathered answers, we first analyze the background questions to provide a descriptive statistics and a contextual overview. By means of (Spearman's) correlation we test the relationship between all questions and statements. We apply Mann-Whitney U and Kruskal-Wallis (for questions with more than two categories) tests to determine any distribution differences in statement agreements across the background questions. A pair-wise comparison of demographic groups with the Mann-Whitney U test is applied (with Bonferroni correction) on significant distribution differences to assess whether the groups agree differently with the statements.

Finally, we compute the mean and standard deviation for the Likert scale answers and rank the statements by computing the proportions of disagreements vs. agreements, visualized centered around zero. We discard the neutral (Neither agree or disagree) answers in the ranking visualization.

**Table 3: Significant Spearman correlations $> 0.5$.**

| Statement 1 | Statement 2 | Corr. | p-value |
|---|---|---|---|
| S11) Ind. fair. (≡) | S12) Ind. fair. (≈) | 0.784 | $< .001$ |
| Q4) Past applic. | Q5) Job biased | 0.582 | $< .001$ |
| Q1) Age | Q2) Job exp. | 0.575 | $< .001$ |
| S10) Pred. par. (PPV) | S12) Ind. fair. (≈) | 0.539 | $< .001$ |

## 4 Outcomes

### 4.1 Participants background

The background questions overview reveals that our participant's age groups concentrate (approx. 60%) between 26 and 45 years old, and 16% and 25% fall into below 25 and above 46 respectively. The majority (63%) reports to have more than 10 years of working experience and 36% participated 1 or 2 times in hiring decision, 28% up to 10 times and 20% not participated; and 16% did so more than 10 times in the last 5 years. In terms of past job applications, 44% applied more than 10 times to a job, 36% 3 to 10 times. More than half (64%) reported to have occasionally experienced some degree of bias during those job applications. When asked whether there is a prevalent type of bias in the current job market in their country, many (23% and 30%) answered that they perceive a bias for age and ethnicity. Interestingly, around 30% of participants provided an open answer ('Other' option) to the latter. The full background questions and answer statistics are in the supplementary material.

Table 3 shows significant, above 0.5 correlations between survey questions where Individual fairness questions (S11 and S12) appear highly correlated, more past applications (Q4) correlate with higher perceived bias in hiring procedures (Q5), age correlates with job experience and Predictive parity (S10) correlates with Individual fairness (S12). The correlation between S11 and S12 suggests the two statements have a very similar meaning, making it worth to use only one for Individual fairness.

The pairwise Mann-Whitney U tests across sub-groups of participants reveal that participants in 36-46 age bracket (15) tended to strongly agree with S3 (Unawareness of gender and age), while participants in 18-25 tended to be more neutral (8) ($U = 21, p = 0.039, \alpha = 0.05$). Similarly, participants who never contributed in hiring (10) and those contributing 1 or 2 times (18) mostly agreed with S5 (Unawareness of location), while those who hired 3-10 times (14) were neutral or strongly disagreed ($U = 114$ and $U = 198$, $p = 0.049$ and $p = 0.028$, $\alpha = 0.05$ respectively).

### 4.2 Evaluating statements

Table 4 shows the means and SD of agreements for all fairness statements, with the top four statements in bold, and Figure 2 shows

**Table 4: Descriptive statistics of user judgments.**

| Statement | Mean | SD |
|---|---|---|
| S1) Diversity goal | 3.69 | 1.05 |
| S2) Demographic parity 1 | 4.06 | 1.17 |
| S3) **Fairness unawareness - gender & age** | **4.24** | **1.01** |
| S4) Fairness unawareness - age | 3.92 | 1.00 |
| S5) Fairness unawareness - address | 3.84 | 1.19 |
| S6) Demographic parity 2 | 2.69 | 1.22 |
| S7) **Equal opportunity (TPR)** | **4.22** | **0.86** |
| S8) **Predictive equality (FPR)** | **4.37** | **0.77** |
| S9) Fairness unawareness - test score | 2.67 | 1.28 |
| S10) Predictive parity (PPV) | 3.98 | 1.05 |
| S11) Individual fairness (≡) | 3.98 | 1.16 |
| S12) **Individual fairness (≈)** | **4.10** | **1.04** |

the ranked agreements grouped by fairness metric. We grouped the answers corresponding to fairness definitions as shown in Table 2, that is, by Demographic parity, Fairness through unawareness, Individual fairness and the remaining three definitions. Neutral answers were excluded from the ranking in Figure 2. The majority of participants expressed strong agreement with all statements except S6 (Dem. parity 2) and S9 (Fair. unawar. - test score). Specifically, more than half of them disagreed with S6 and S9.

The first disagreement (S6) highlights a contrast of participants preferences for statements representing Demographic parity (S2 and S6): whereas participants mostly agree with the S2 formulation of Demographic parity, almost half of them shows disagreement with the specific formulation with unequal base rates of male and female candidates. While users overwhelmingly agree with excluding gender (S3), age (S4) and address (S5) attributes in the hiring decision, the disagreement with S9 suggests that neither it can depend only on the test score .

The user disagreement with statements S6 and S9 indicates that such statements might not be the best option to use for collecting user preferences on these fairness requirements (Demographic parity with unequal base rates, and Fairness through unawareness of a single objective attribute). Therefore, the corresponding requirements would fall short on perceived fairness and are not recommended to include in a requirement elicitation discussion.

## 4.3 Fixing concrete fairness requirements

Following Figure 2, the highest ranked statements - S8) Predictive equality (FPR) and S7) Equal opportunity (TPR) - indicate users' preference for same chances to be *misclassified* for unqualified candidates, and same chances to be classified correctly when truly qualified. Fairness through unawareness of gender and age (S3) ranks third, and approximate Individual fairness (S12) ranks fourth.

With the obtained preferences from the users, we can prioritize our fairness requirements and translate the best ranked metrics into concrete specifications. For example, since S8) Predictive equality (FPR) is the highest ranked metric, one can specify as a software requirement for the ADM system to prioritize training on balanced data between the protected and unprotected groups to avoid one group to over-predict positives, at the cost of accuracy [5]. We can set more than one fairness specification if they are compatible,

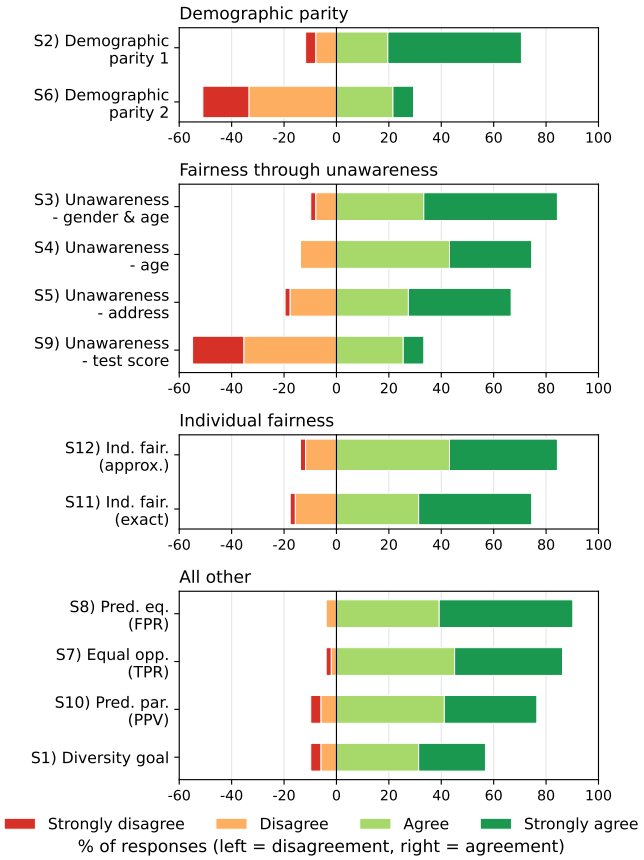



**Figure 2: Agreements grouped by fairness definitions.**

such as including the second ranked fairness definition - S7) Equal opportunity (TPR). The two metrics are indeed typically compatible (Table 1), this means that we can specify the classification algorithm to implement mitigation techniques for both, such as re-balancing training data and/or by adjusting classification thresholds [22].

While Figure 2 shows that most participants agree with almost all formulations of the six fairness definitions, recall the practical trade-off in Section 2.2 where the fairness metrics are - in practice - seldom compatible with each other. Decision-makers and developers need to decide on the trade-off mentioned in Section 2.2 of which fairness requirements are preferred and which can be implemented. We discuss our pragmatic approach to settle on a given fairness requirement in Section 5.

## 4.4 Open answers analysis

In Q6 ("*Which type of bias do you think is most prevalent in the job market in the county you currently live in?*"), eight participants expressed their perceived bias as related to attributes like residence, education, provenance and "*old boys network*" - "*the need to know somebody in the company*". Interestingly, two participants mention specifically the arbitrariness of automatic hiring systems and the lack of technical knowledge of some HR professionals as a potential risk. This emphasized the need for training of people operating AI systems, as required by the AI Act [14].

The final question of the survey produced useful insights into what people consider important in this application. Multiple participants mentioned the importance of designing a good process around the use of decision algorithms, based on flexible criteria (e.g., no hard exclusions based on experience), with multiple rounds and not showing names and photos, or other irrelevant features to people selecting candidates. Other interesting suggestions are to look at changes in the audition process of musicians for inspiration, to provide proper training and raise awareness in companies about potential bias, and the need for carefully written requirements for a position: an irrelevant requirement needlessly excludes candidates. We recommend algorithm designers to ask an open question to stakeholders for input on the overall recruitment process.

### 4.5 Limitations

Our study is limited in both geographical reach (most participants from Western Europe) and in background information collected (no gender or country of origin). This is a potential area for further research. Another limitation concerns the size and opportunistic nature of the sample, with outcomes not generalizable to a wide population. We therefore do not recommend people to rely on our outcomes. Instead, algorithm implementers should deploy their own survey among their stakeholders. Similarly, our scenario (and the example hiring dataset) is limited to a specific hiring case and few attributes. Companies should analyze their requirements and determine which features are sensitive in their case (we consider multiple sensitive attributes, including but not limited to gender).

As each fairness definition was operationalized by a single representative statement to identify the most 'agreeable' formulations and to keep the survey pragmatically short for practical use, internal consistency metrics (e.g., Cronbach's $\alpha$) were not applicable. Instead, we focused on construct validity through derivation from formal definitions, expert review and pilot testing. We can only assess construct validity by examining the discriminant validity across unrelated fairness statements, which showed no strong correlations between them (only Individual fairness statements showed a high correlation in Table 3, and thus can be removed in the future). Still, the findings are limited by potential variations in how respondents interpret fairness terminology. Future work could strengthen reliability and validity through comprehension checks, multi-item constructs and factor analysis, and replication beyond the hiring domain. Finally, our discussion of fairness trade-offs is conceptual in nature, and future research efforts should strive to ground the fairness trade-offs in observation and experience.

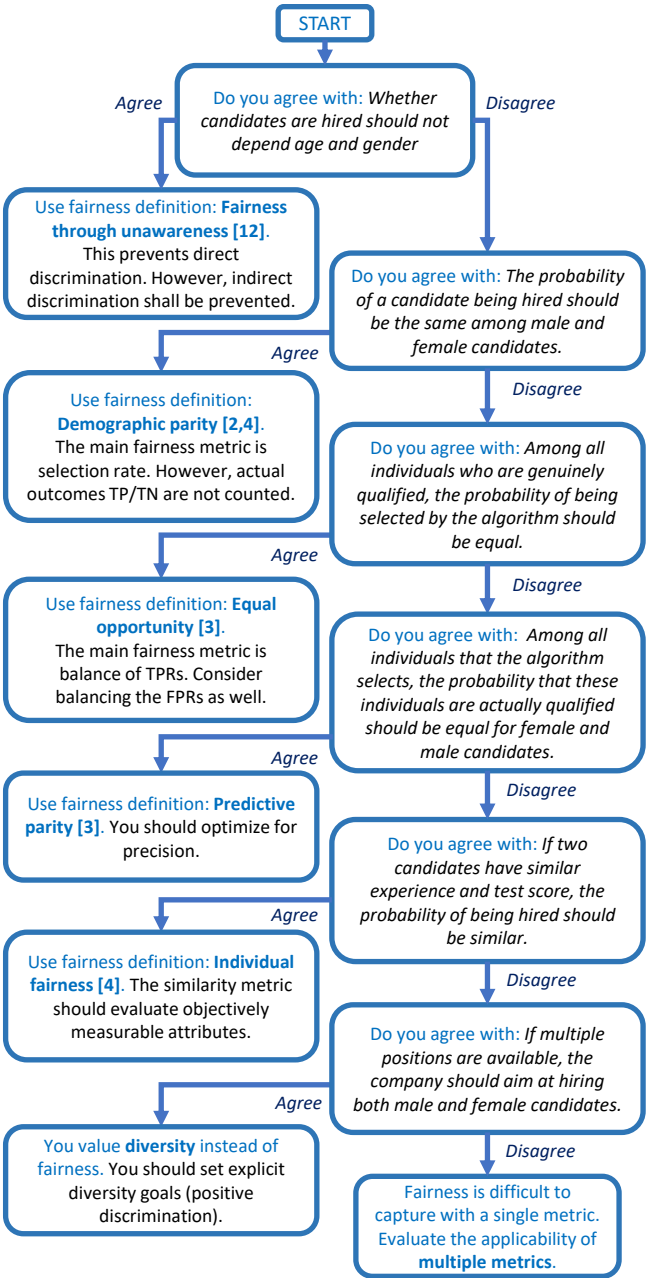



**Figure 3: Conceptual fairness requirement decision tree.**

## 5 Discussion and Conclusion

The analysis of the survey outcomes shows that our survey includes relevant statements that can be used to determine fairness requirements. Most participants agree with several statements (e.g., S7, S3) and do not report any issues in the open question. This research thus shows how fairness requirements can be collected. Based on the evaluation, it is possible to improve the survey by combining statements that are highly correlated, that is, S11) and S12). Similarly, S6) and S9) were negatively perceived (as not fair) by most participants and are thus less important to include.

Having evaluated the statement formulations with which users agree the most, we summarize in Figure 3 a conceptual infographic of statements that practitioners can use to gauge stakeholder preferences and guide discussions around resolving the fairness requirements trade-off. The infographic can be used as a starting point for a risk assessment and a requirements interview. It can also be used to generate awareness on the need to carefully evaluate AI applications and educate people on the challenges of achieving fair decision making.

## References


[1] Pavlo Burda and Sieuwert Van Otterloo. 2025. Fairness definitions explained and illustrated with examples. *Com. Soc. Res. J.* 2 (Aug. 2025), 1–22. doi:10.54822/PASR6281

[2] Alessandro Castelnovo, Riccardo Crupi, Greta Greco, Daniele Regoli, Ilaria Giuseppina Penco, and Andrea Claudio Cosentini. 2022. A clarification of the nuances in the fairness metrics landscape. *Sci Rep* 12, 1 (March 2022), 4209. doi:10.1038/s41598-022-07939-1 Publisher: Nature Publishing Group.

[3] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (June 2017), 153–163. doi:10.1089/big.2016.0047 Publisher: Mary Ann Liebert, Inc., publishers.

[4] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12)*. ACM, New York, NY, USA, 214–226. doi:10.1145/2090236.2090255

[5] Alessandro Fabris, Nina Baranowska, Matthew J. Dennis, David Graus, Philipp Hacker, Jorge Saldivar, Frederik Zuiderveen Borgesius, and Asia J. Biega. 2025. Fairness and Bias in Algorithmic Hiring: a Multidisciplinary Survey. *ACM Trans. Intell. Syst. Technol.* 16, 1 (Feb. 2025), 1–54. doi:10.1145/3696457 arXiv:2309.13933.

[6] Batya Friedman, Peter H Kahn, and Alan Borning. 2002. *Value Sensitive Design: Theory and Methods*. Technical Report. University of Washington.

[7] Nina Grgic-Hlaca, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. Int. World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 903–912. doi:10.1145/3178876.3186138

[8] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*, Vol. 29. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

[9] Galen Harrison, Julia Hanson, Christine Jacinto, Julio Ramirez, and Blase Ur. 2020. An empirical study on the perceived fairness of realistic, imperfect machine learning models. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. ACM, New York, NY, USA, 392–402. doi:10.1145/3351095.3372831

[10] Styliani Kleanthous, Maria Kasinidou, Pınar Barlas, and Jahna Otterbacher. 2022. Perception of fairness in algorithmic decisions: Future developers' perspective. *PATTER* 3, 1 (Jan. 2022). doi:10.1016/j.patter.2021.100380 Publisher: Elsevier.

[11] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2016. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807* (2016).

[12] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual Fairness. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

[13] Stefan Leijnen, Henry Maathuis, Kees van Montfort, Sieuwert Van Otterloo, Danielle Sent, Marcel Stalenhoef, Koen van Turnhout, and Raymond Zwaal. 2025. *Guideline for user interface design of explainable AI*. Technical Report. Hogeschool Utrecht. https://www.hu.nl/onderzoek/publicaties/guideline-for-user-interface-design-of-explainable-ai

[14] European Parliament. 2023. EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[15] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q Weinberger. 2017. On Fairness and Calibration. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2017/hash/b8b9c74ac526fffbeb2d39ab038d1cd7-Abstract.html

[16] Elissa M Redmiles, Yasemin Aca, Sascha Fahl, and Michelle L. Mazurek. 2017. *A Summary of Survey Methodology Best Practices for Security and Privacy Researchers*. Technical Report. Computer Science Department CS-TR-5055. University of Maryland.

[17] Carlotta Rigotti and Eduard Fosch-Villaronga. 2024. Fairness, AI & recruitment. *Computer Law & Security Review* 53 (July 2024), 105966. doi:10.1016/j.clsr.2024.105966

[18] Megha Srivastava, Hoda Heidari, and Andreas Krause. 2019. Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, New York, NY, USA, 2459–2468. doi:10.1145/3292500.3330664

[19] Christopher Starke, Janine Baleis, Birte Keller, and Frank Marcinkowski. 2022. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society* 9, 2 (July 2022), 20539517221115189. doi:10.1177/20539517221115189 Publisher: SAGE Publications Ltd.

[20] Saurabh Tiwari, Santosh Singh Rathore, and Atul Gupta. 2012. Selecting requirement elicitation techniques for software projects. In *2012 CSI Sixth International Conference on Software Engineering (CONSEG)*. IEEE, Indore, Madhay Pradesh, India, 1–10. doi:10.1109/CONSEG.2012.6349486

[21] Sahil Verma and Julia Rubin. 2018. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness (FairWare '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3194770.3194776

[22] Hilde Jacoba Petronella Weerts. 2025. *Decoding Algorithmic Fairness: Towards Interdisciplinary Understanding of Fairness and Discrimination in Algorithmic Decision-Making*. Phd Thesis 1 (Research TU/e / Graduation TU/e). Eindhoven University of Technology, Eindhoven. https://research.tue.nl/en/publications/decoding-algorithmic-fairness-towards-interdisciplinary-understan/ ISBN: 9789038663357.