Shifting Paradigms: Value Sensitive Design for Fair AI Recruitment

Alexandre Puttick[1], Carlotta Rigotti[2], Ahmed Abouzeid[3], Eduard Fosch-Villaronga[2], Mascha Kurpicz-Briki[1] and Pinar Øztürk[3]

[1]*Berner Fachhochschule BFH, Technik und Informatik, Quellgasse 21, 2501 Biel, Switzerland*

[2]*Leiden University, Rapenburg 70, 2311 EZ Leiden, Netherlands*

[3]*Norwegian University of Science and Technology, IT-bygget, almen, Sem Sælands vei 9, 7034 Trondheim, Norway*

Abstract

In this position paper, we advocate for the use of *value sensitive design* (VSD) as a framework for developing *fair AI recruitment tools*. As a starting point, we assert that the current paradigm in AI fairness in the hiring context is severely limiting. We then document an ongoing process within the EU-horizon project BIAS, seeking to escape this paradigm by applying VSD to the development of AI applications for candidate selection with diversity and fairness as focal points. In particular, we present *case-based reasoning* as a case study in the intentional operationalization of stakeholder positions on fairness and detail how such an approach can be further expanded, drawing from the concept of *agonistic machine learning*. In this endeavor, we hope to contribute to the discourse on the ethical design and use of AI within the labor market and in general.

Keywords

AI, fairness, value sensitive design, recruitment, diversity bias

Introduction

The last decades have seen a growing trend toward designing and deploying artificial intelligence (AI) applications for recruitment and selection. However, such tools lack transparency and pose the risk of algorithmic diversity bias,[1] reinforcing harmful stereotypes and power structures, and, on the individual level, acting to the detriment of dignity, autonomy, and well-being. Consequently, job candidate profiling is classified as *high risk* under the EU AI Act.[2]

As a starting point for this position paper, we assert that the current paradigm in AI fairness in the context of recruitment is severely limiting. Attempts to promote fairness in AI hiring tools typically take diversity metrics as a starting point and modify the training or computational aspects of existing ranking/scoring models to improve metric performance. This approach

AIMMES 2025 Workshop on AI bias: Measurements, Mitigation, Explanation Strategies | co-located with EU Fairness Cluster Conference 2025, Barcelona, Spain

alexandre.puttick@bfh.ch (A. Puttick); c.rigotti@law.leidenuniv.nl (C. Rigotti); ahmed.abouzeid@ntnu.no (A. Abouzeid); e.fosch.villaronga@law.leidenuniv.nl (E. Fosch-Villaronga); mascha.kurpicz@bfh.ch (M. Kurpicz-Briki); pinar@ntnu.no (P. Øztürk)

[1]Unfair positive or negative treatment of individuals, primarily based on protected grounds, thereby aligning the concept with established legal understandings of discrimination, including sex, race, color, ethnic or social origin, genetics, language, religion or belief...

[2]Annex III: High-Risk AI Systems Referred to in Article 6(2), Point 4.

bypasses the explicit consideration of values and normative assumptions and fails to engage with the question of whether the technical interventions and metrics are truly aligned with ethical fairness goals. It operates within the framework of a rigid, deterministic technical solution whose blueprint implicitly assumes a set of normative stances that can hinder fairness while precluding the innovation of novel technical approaches. The resulting models remain opaque to users and impose a predetermined balance of values, stifling moral agency and accountability and leading to unjustifiable decisions.

This paper documents an ongoing process within the EU-horizon project BIAS, seeking to escape this paradigm by applying Value Sensitive Design to the development of AI applications for candidate selection with attention to diversity and fairness. Section 1 introduces the necessary background aspects of VSD. A description of the application of VSD towards fair AI recruitment tools follows in Section 2, laying the foundation for the pro-active technical approaches described in Section 3.

1. Value Sensitive Design

Value sensitive design (VSD) was developed by Batya Friedman and Peter Kahn in order to promote alternative design goals and methods built upon respect for human agency and responsible computing [1]. The theory assumes that every technology is imbued with and reproduces particular human values. *There is no such thing as objective, value-neutral technology.* As such, VSD promotes proactive engagement with human values throughout the design process, aiming to spur technological innovation that is consciously driven by values that aim to improve the human and planetary condition. As a starting point, *human values* are anything that a person or group of people consider highly important. The open-ended understanding of human values and the resistance to reifying a particular set of morals is seen as a strength by proponents of VSD; the emphasis on recognizing co-existing, often competing human values and the need to carefully and intentionally balance them counters the misguided idea that there exists a singular solution that is best for everyone. As opposed to developing a technology according to the values of its designers, VSD insists on in-depth investigation of stakeholder values and how they relate to each other.

Values, Normative Assumptions and Technical Methods. In practice, the development of new technology under the VSD framework involves the identification of the values important to each group of primary stakeholders (e.g. *fairness*) and the corresponding *normative assumptions/stances* about what those values mean (e.g. *"Everyone should be treated equally, regardless of their background," or "People with facing greater structural inequality should receive more resources to ensure equitable outcomes."*). Finally, an investigation of technical methods that operationalize these normative stances (e.g. *The requirement to meet certain fairness metric thresholds.*) is conducted, with careful consideration of the balancing and/or reconciliation of competing values. This should all be considered within the sociotechnical context in which the technology is situated, the relationships it mediates, and the social effects its use could have over time.

Tri-Partite Methodology. VSD recommends a methodology consisting of *conceptual, empirical* and *technical* investigation. At the conceptual level, researchers are expected to consider the specific values and normative assumptions in which a technology is situated, clarifying fundamental issues raised by the technology and design process. It focuses on theoretical considerations of values, potential conflicts, and trade-offs. During empirical investigations, researchers position the design within the relevant social context, seeking to identify, e.g., important stakeholders and understand their values. One gathers data from real-world stakeholders to understand their experiences, needs, and concerns. Technical investigations aim to identify specific technological approaches suited to supporting certain values, with attention to potential detriment to others. This includes assessing the capabilities and limitations of technical solutions. Friedman highlights two subtypes of technical investigation. The first, *retrospective analyses*, focuses on "how existing or historical technological processes and underlying mechanisms support or hinder human values," [2]. The second subtype, *proactive design*, concerns the design of systems that support the values and normative stances identified during conceptual and empirical investigations.

2. Fairness in AI Recruitment

Values and Direct Stakeholders

The primary values of interest are *diversity, fairness* and the *selection of the most capable job candidate*. We concentrate on two sub-types of fairness: 1) *Procedural*, concerning how the decision was made; 2) *Substantive*, concerning the outcome of the decision. VSD defines *stakeholders* as those who are or will be significantly implicated by the technology [2]. The primary direct stakeholder groups identified within the BIAS project are *AI developers/researchers, HR professionals, workers/job seekers* and *human rights advocates/regulators.* Since fairness and diversity are central values within the project, emphasis is placed on the representation of underprivileged identities within each stakeholder group.

Tri-Partite Investigation of Fairness

Conceptual Investigation

Defining Fairness. In the broad social and EU legal context, *substantive fairness* in relation to diversity may be viewed in terms of *group fairness* with respect to hiring outcomes, i.e., the fair distribution of jobs amongst demographic groups [3]. *Procedural fairness* emphasizes a *well-grounded, transparent decision-making process in which discrimination on the basis of protected attributes is restricted.* Based on a literature review and empirical research, we have identified the following stakeholder perspectives:

Job applicants: Substantive fairness is perceived in terms of hiring decisions that reflect their knowledge, skills, and efforts [4, 5, 6, 7, 8]. Also, applicants are concerned with procedural fairness, e.g., in terms of 1) *job relatedness*: The selection process should only assess the personal characteristics that are necessary for the job; 2) *consistency*: Each applicant should go through

the same process; and 3) *opportunity to perform*: The applicant can demonstrate their knowledge and skills during the hiring process; [9, 4, 10, 11, 12, 13, 14, 15, 16, 17].

HR practitioner/company: The decision-making process should be aimed toward selecting the most capable candidate for the job [18, 19, 20]. At the team level, diversity is desirable but must be balanced against the company's community/cultural structures [21].

Empirical Investigation

One of our main methodologies for empirical investigation consists of co-creation workshops carried out in all nine partner countries. These workshops focus on understanding the views of different stakeholders, as well as their desires, needs and opinions regarding what constitutes a fair and useful recruitment tool, making use of mock tools[3] to stimulate discussion. *Agency* and *transparency* in relation to both *users* and *job applicants*, as well as *technical robustness*, were highlighted as key aspects of a fair and trustworthy decision-making process.[4] The outcomes of the decision-making process–algorithmic, human or hybrid–*are not fair if they are not justifiable.* Participants brought up situations exemplifying unfairness, such as *candidate scores being artificially raised* by invisibly pasting job ads into their applications or *being rejected based on incorrectly parsed data.* A *disconnect between how outliers are perceived by statistical models and the desirability of candidates with unique profiles* also arose as a potential source of unfairness. There was additional concern over how *the omission of particular types of candidates by a deterministic model becomes systemic, whereas multiple different evaluators may compensate for each other's particular biases.*

Technical Investigation - Retrospective Analysis

Encoded Values and Hindrances to Fairness. Compared to human decision-making, existing tools claim to save time and money for employers while simultaneously making the recruitment process more objective and accurate. VSD rejects the notion that technology can ever be objective; terms like *objectivity* and *accuracy* mask normative assumptions behind neutral-sounding terminology, and the repeated discovery of encoded bias in machine learning models has made it clear that such technologies are not objective. This is often explained via a *garbage in, garbage out* understanding of bias in AI systems, but it is important to note that *value-encoding goes beyond training data.* For example, ostensibly neutral metrics such as *loss* and *accuracy scores* implicitly reinforce the values underpinning the status quo, while predictive modeling treats unclassifiable outliers as data points to be ignored.

The fundamental task–using a pool of candidate data, compute the best candidates—assumes that *job suitability* is a one-dimensional trait, largely based on a pre-defined notion of what constitutes a high-achieving individual. It is associated with *objectivity* because suitability is usually evaluated based on educational background, work experience and other "objective"

[3]These simulated LLM-driven AI tools, allowing users to specify and weight various criteria, providing natural language justifications for candidate rankings, and flagging potential unfair biases.

[4]User agency over the criteria under which candidates are evaluated, the ability to flag system errors or potential harmful behaviors, an interactive process that allows users to explore candidate data from multiple quantitative and qualitative angles, and the capacity for the tool to draw attention to their own unconscious biases.

criteria. But this ignores structural inequalities that can render apparently objective features into grounds for *proxy/indirect discrimination*. By collapsing each candidate to a small subset of features to be evaluated by a fixed algorithm, job suitability scores can work to the detriment of creating a diverse employee pool with complementary skills.

Fairness Methods. The technical research community has predominantly focused on operationalizing *group fairness* [3]. The most widely adopted *fairness metrics* are thus derived from demographic distributions in algorithmic output. As described at length in [22], these metrics are (very rough) proxies for often implicit normative stances. *Good performance on fairness metrics is no guarantee of fairness*, and technical methods to optimize fairness metrics do not explicitly engage with underlying normative stances and are hence not necessarily fairer. Furthermore, recent work demonstrates that no established mathematical notion of fairness sufficiently captures stakeholder notions of fair hiring practice [23]. Fairness metrics are thus *valuable in demonstrating model unfairness or evaluating fairness interventions that are made in line with ethical values, but should not be the targets of blind optimization nor utilized as proofs of fairness.*

The Fairness/Accuracy Trade-Off. Research shows that there are unavoidable trade-offs in attempting to simultaneously optimize performance and fairness metrics [24]. Through the lens of VSD, this can be viewed as a numerical manifestation of value tensions and could be reframed as a *social justice/status quo* trade-off or a *diversity/job suitability* trade-off. However, it should be noted that VSD warns against framing the balance of values in terms of trade-offs; doing so predisposes the designer to seek approaches in which promoting one value will diminish another, rather than seeking ways to promote both, e.g., by *changing the evaluation criteria for job suitability* or *scoring candidates in a way that is less entangled with the labels present in historical hiring data.*

3. Proactive Design

Case-Based Reasoning (CBR) and Fairness

Case-Based Reasoning (CBR) forms a foundational AI approach for our proposed Decision Support System (DSS) in recruitment. This system assists HR professionals during the candidate screening process and produces one of three potential outcomes for a candidate applying for a specific job opening: *"Shortlist," "Longlist,"* or *"Negative."* In a CBR engine, the *case-base* consists of a database of past candidate data along with the decision that was reached on each candidate's application. The system compares new candidates to the case-base, retrieves past similar candidates, and a decision is reached by aggregating those made in the past. The *features stored in the case-base* and the *similarity metric* are two key points where the designer has direct influence over the operationalization of fairness principles, including the integration of *domain knowledge*; the model's computations are grounded in the decision-making processes of HR professionals. The CBR workflow is depicted in Figure 1.

We emphasize that the development of the CBR engine is fully situated within the intended deployment context; the data and design insights were gathered from our industry partner,

and the model is bespoke for their own needs and internal processes. At the same time, the model framework, principles and parameters can be readily adapted to new industry contexts. We assert that this is preferable over the development of a generic tool targeting the broadest possible context; one-size-fits-all approaches impose natural limits to *accountability, transparency, justified decision-making processes, user agency* and *the use of domain knowledge.*

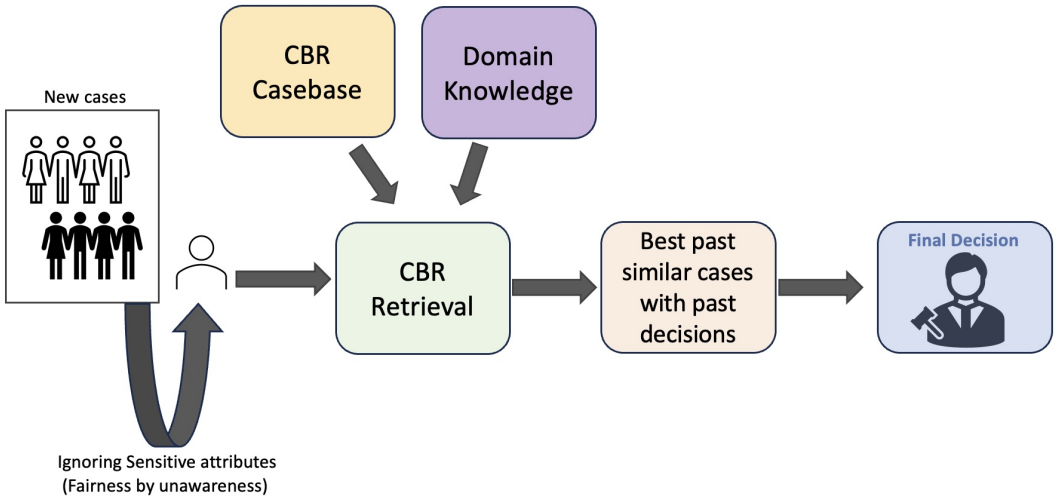

Figure 1: CBR Pipeline

Data. The CBR system depends on careful *feature engineering* to determine precisely the information on which the model bases its decisions. To effectively capture domain-specific insights, our approach combined workshops and surveys with HR professionals from our industry partner. This mixed methodology collects both quantitative and qualitative observations about the candidate data and the operational context of the recruiting company. The domain insight we gather represents HR experts' background knowledge about the job types, attributes, and the relevant range of values for those attributes[5] for the given sector and company. Another key aspect of our system is capturing the relative importance of each attribute in the decision-making process. The importance of an attribute is context-sensitive and can vary across different job types. To ensure the validity of our findings, it was crucial to gather responses from multiple HR personnel employed by our industry partner. This approach helps mitigate potential biases, as different evaluators may interpret applicant data in distinct ways. Figure 2 demonstrates the extracted general attribute importance from the survey results.

Model. The importance measured for each feature determines a baseline for their relative weight in computing similarity scores, which can be further adjusted according to domain. This

[5]E.g. *Bachelor's* or *Master's* for *Educational Level*, as opposed to *Some higher education* and *No higher education.*

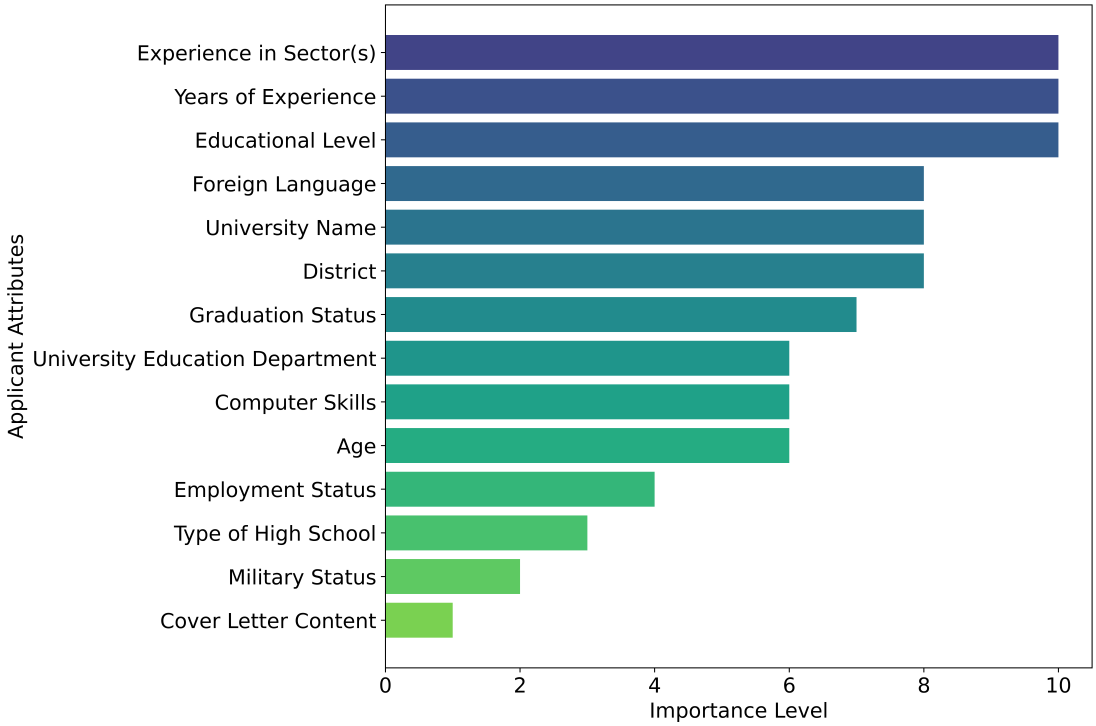

Figure 2: Survey results depicting the general importance of attributes for recruitment decisions.

is represented by the equation

$$S_{\text{global}}(X, Y) = \sum_{i=1}^{n} w_i \cdot s_{\text{local},i}(x_i, y_i), \tag{1}$$

Where X and Y consist of the feature vectors of two candidates, and w_i and $s_{\text{local},i}$ respectively denote the *weight* and *job-specific similarity metric* attributed to the ith feature. For each new candidate, the most similar candidates from the case-base are retrieved by computing similarities using the extracted, engineered features and relative weights. A decision is then rendered by aggregating the decisions made on the retrieved cases, weighted by similarity.

Evaluation. A key element of procedural fairness is that every candidate is subjected to the same process. A weakness of the CBR engine is that its reliability depends on having sufficiently many similar candidates within the case-base. This is essentially a version of the outlier problem that plagues all statistical models. We thus define the decision *certainty* as a metric for *procedural/individual* fairness; it is inversely proportional to the average distance between a new case and the retrieved cases, as computed using the similarity metric in Eq. 1. As opposed to machine learning, one can use certainty to immediately if the candidate in question is an outlier, so that their application may be considered with further scrutiny rather than systematically rejected, and the case-base can be immediately updated. Whereas a machine

learning model may require thousands of such updates to perform reliably on future, similar candidates, a CBR-engine only requires a number proportional to the number of cases retrieved for decision-making ($N \approx 10$).

Operationalizing Fairness. The setup we describe attempts to operationalize several fairness notions established during our investigations:

- Omitting any direct information about protected attributes prevents *direct discrimination* and limits the possibility of *proxy discrimination* compared to machine-learning-based methods using raw text data.
- The selection of features deemed highly relevant in the candidate selection process and weighted according to user specification promotes *job relatedness, user agency*, and *selection of the most qualified candidate* in a *well-grounded and explainable decision-making process*.
- The decision to sort candidates into rough categories instead of rankings acknowledges that job suitability is not one-dimensional.
- The certainty metric provides a proxy for *robustness* and *reliability*, aspects of *procedural fairness*, as well as the capability to flag outliers who may otherwise be treated unfairly.
- The ability to rapidly update the case-base, and customizability of extracted features and similarity scores further support technical robustness and post-deployment adaptivity, acting as guardrails against the hegemony of static algorithms.

Beyond CBR

The CBR engine represents a design framework and prototype developed with explicit attention to stakeholder perspectives on fairness, but we acknowledge its shortcomings; the engine's output is still dependent on historical hiring decisions that may be unfair, while features rated highly by HR professionals such as University Name, District, Foreign Language and Employment Status could serve as proxies for protected attributes and be used to reinforce unfair biases in the hiring system rather than mitigate them. We thus conclude this position paper with a summary of further research directions in proactive design that can complement CBR and further integrate stakeholder fairness principles.

(Large) Language Models

Deep-learning-based language models extract complex features from raw text data, achieving unparalleled performance and completing tasks that would be intractable with simpler, interpretable models such as the CBR engine. The language comprehension capabilities of large language models (LLMs) such as ChatGPT and DeepSeek further expand the frontier of possibilities. Such models can offer *qualitative analyses* that are not possible using numerical methods and could further promote fairness. However, what these models gain in capability is paid for with opaque processes that hinder transparency, agency, and justifiability. Moreover, a large body of research demonstrates that such models encode harmful stereotypes that can lead to negative outcomes when integrated into AI recruitment tools (e.g., [25]).

Bias Detection and Mitigation. In parallel to the CBR engine, the second main line of technical research in the BIAS Project consists of a framework for and collection of bias detection and mitigation techniques for various stages along the AI recruitment pipeline.

- Any recruitment tool that claims not to consider protected traits in the decision-making process should be able to demonstrate that those traits cannot be inferred from training or input data. We train a series of protected attribute classifier models for benchmark testing: *In the absence of proxies for protected attributes, such classifiers should not perform better than random chance.*
- We are developing a large repository of bias detection and mitigation tools for word embeddings and language models, focusing on *EU languages in the occupational context*. These are intended to help developers make a concentrated effort to mitigate harmful stereotypes encoded in AI language models used in the recruitment context.
- A suite of tools for flagging potential grounds for unfair bias in candidate applications is also being developed. On the candidate side, these tools could suggest rewrites to help the candidate avoid prejudice, while, on the user side, they can either flag for consideration or mask such information from the decision-making process.

Justifying Decisions. Following [26], we understand *justifications* in the context of decision-making AI to mean "not merely explaining the logic and the reasoning behind [automated decisions], but also explaining why it operates in a legally acceptable (correct, lawful, fair) way." This is in contrast to *explanations* in the sense of *explainable AI* (XAI), which seek to shed light on the internal decision-making process of the model. Justifications are often not only more feasible, but more useful and desirable; they can serve as accountability measures for AI-aided decision-making and address, e.g., the right to an individual explanation or contestation of algorithmic decisions. Using LLMs, we have experimented with the generation of justifications that cite elements of job postings and cover letters to support the decision rendered on a specific applicant, while also referring to provisions in EU anti-discrimination law to justify decisions made in the interest of substantive equality/positive discrimination. These justifications have been utilized in mock tools to gather stakeholder feedback.

Agonistic Machine Learning

The fairness perspectives described in this work can be complementary or adversarial, and VSD asserts that balancing these values must be conducted with care and intention. In the context of AI tools, a candidate's data stands as a proxy for the candidate themselves, and ranking algorithms are typically calibrated to a particular, implicit balance of values, processing candidate data and returning a deterministic output. Kate Crawford [27] and Mireille Hildebrandt [28] find fault in this framing and propose an alternative framework called *agonistic machine learning*. Agonistic machine learning is built upon the notion of the *incomputable self*, that "any computation of our interactions can be performed in multiple ways–leading to a plurality of potential identities" [28]. There is no single correct way to represent a candidate by data, to process that data, nor to present the result of that processing—no best, most fair ranking. The term *agonistic* refers to collective decision-making processes involving choices between

conflicting options–not necessarily by achieving rational consensus, but through a struggle between adversaries.

Agonistic Recruitment Tools. As an ideal, we imagine a *multi-model, multi-human decision-making process* in which the aim of an AI recruitment tool is not simply to generate a single ranked shortlist of candidates, but rather to *facilitate a deliberative process by presenting candidates according to multiple criteria and various value-balances*. A CBR engine, bias-mitigated language models and qualitative LLM components can all be combined in a multi-model agonistic ecosystem, which can be further expanded to include fairness methods such as fair ranking and counterfactual data augmentation. Models should incorporate *randomness* or aim to *identify outliers* as means to avoid unfair systematic exclusion and place value on uniqueness. Such an ecosystem extends the notion of *ensemble models*, in which the output and capabilities of multiple models are aggregated to balance the strengths and weaknesses of individual components, e.g., combining simpler models (CBR) with black-box neural networks to enable a fine-tuned balance between interpretability and performance. By explicitly establishing numerical proxies for different normative stances, mathematical constructs such as the *Pareto front* can be used to select from a set of aggregating schemes that minimize value trade-offs. Moreover, there are existing guidelines for exploring the space of possible models according to various value balances [29].

Conclusion

This work represents a reversal of a status quo in which fairness metrics act as proxies for undefined normative stances and existing models are modified to optimize them. Deep engagement with stakeholder values should precede technical methods, which can then be built from the ground up to operationalize and balance those values. To summarize our contributions,

- An explicit set of values and stakeholders is determined and the tri-partite investigation of stakeholder normative stances and existing technical methods is described.
- The case-based reasoning engine is presented in the context of pro-active design, with explicit attention towards the operationalization of stakeholder fairness perspectives.
- The CBR engine is situated within a larger technical research and development environment that engages with fairness aspects that are not fully encompassed within the CBR framework.
- We draw from the notion of agonistic machine learning as a means to combine and balance the values underlying individual models and evaluation methods while further promoting moral agency and fully engaging with the fact that candidates cannot be reduced to a particular set of numerical features or a single score or deterministic automated decision.

VSD emphasizes *progress, not perfection*; the ideas described in this article aim to stimulate further discussion and research.[6] Designers bear a moral responsibility for the values embedded

[6]An EU Horizon sister project, FINDHR, has recently published parallel work detailing their design of fair recruitment systems based on VSD [30], further supporting the position presented here. The topic of VSD was discussed between members of our respective projects, but work was conducted independently.

within the technologies they develop. The role of AI in society is rapidly expanding, while ethical, regulatory and social perspectives on AI, DEI initiatives and social justice as a whole are in a state of flux and turmoil. We are at a critical point in history where profound engagement with the interplay between human values and evolving sociotechnical contexts is of crucial importance.[7]

4. Acknowledgements

This work is part of the Europe Horizon project BIAS, grant agreement number 101070468, funded by the European Commission, and has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

References

[1] B. Friedman, P. H. Kahn, Human agency and responsible computing: Implications for computer system design, Journal of Systems and Software 17 (1992) 7–14. URL: https://linkinghub.elsevier.com/retrieve/pii/016412129290075U. doi:10.1016/0164-1212(92)90075-U.

[2] B. Friedman, D. Hendry, Value sensitive design: shaping technology with moral imagination, The MIT Press, Cambridge, Massachusetts, 2019.

[3] S. Wachter, The Theory of Artificial Immutability: Protecting Algorithmic Groups Under Anti-Discrimination Law (2022). URL: https://arxiv.org/abs/2205.01166. doi:10.48550/ARXIV.2205.01166, publisher: arXiv Version Number: 1.

[4] S. W. Gilliland, The Perceived Fairness of Selection Systems: An Organizational Justice Perspective, The Academy of Management Review 18 (1993) 694. URL: http://www.jstor.org/stable/258595?origin=crossref. doi:10.2307/258595.

[5] T. J. Thorsteinson, A. M. Ryan, The Effect of Selection Ratio on Perceptions of the Fairness of a Selection Test Battery, International Journal of Selection and Assessment 5 (1997) 159–168. URL: https://onlinelibrary.wiley.com/doi/10.1111/1468-2389.00056. doi:10.1111/1468-2389.00056.

[6] L. D. Zibarras, F. Patterson, The Role of Job Relatedness and Self-efficacy in Applicant Perceptions of Fairness in a High-stakes Selection Setting: Selection Fairness Field Study, International Journal of Selection and Assessment 23 (2015) 332–344. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12118. doi:10.1111/ijsa.12118.

[7] Recent political developments demonizing DEI efforts appear to have already influenced ChatGPT. An inquiry with no prior context, dated 12.2.2025, asked, "*What is the definition of diversity bias?*" and received the response, "*Diversity bias refers to a type of bias that occurs when efforts to promote diversity lead to unintended discrimination, favoritism, or misrepresentation. It can manifest in several ways, such as:...Tokenism – When organizations include individuals from diverse backgrounds for appearance's sake rather than fostering genuine inclusion. Reverse Discrimination – When attempts to promote diversity result in bias against traditionally dominant groups. Diversity bias can occur in hiring, media representation, education, and decision-making processes...*" The ripple effects that polarized politics and the political leanings of AI tech giants could have on AI tools in general–let alone in recruitment–are both harrowing and difficult to quantify.

[7] S. Schinkel, A. Van Vianen, D. Van Dierendonck, Selection Fairness and Outcomes: A field study of interactive effects on applicant reactions: Selection Fairness and Outcomes, International Journal of Selection and Assessment 21 (2013) 22–31. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12014. doi:10.1111/ijsa.12014.

[8] A. Köchling, M. C. Wehner, Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses, International Journal of Selection and Assessment 31 (2023) 45–62. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12412. doi:10.1111/ijsa.12412.

[9] Annelies E. M. Van Vianen, Ruben Taris, Eveline Scholten, Sonja Schinkel, Perceived Fairness in Personnel Selection: Determinants andOutcomes in Different Stages of the Assessment Procedure, International Journal of Selection and Assessment 12 (2004) 149–159.

[10] S. W. Gilliland, M. Groth, R. C. Baker, A. E. Dew, L. M. Polly, J. C. Langdon, IMPROVING APPLICANTS' REACTIONS TO REJECTION LETTERS: AN APPLICATION OF FAIRNESS THEORY, Personnel Psychology 54 (2001) 669–703. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1744-6570.2001.tb00227.x. doi:10.1111/j.1744-6570.2001.tb00227.x.

[11] D. D. Steiner, S. W. Gilliland, Procedural Justice in Personnel Selection: International and Cross-Cultural Perspectives, International Journal of Selection and Assessment 9 (2001) 124–137. URL: https://onlinelibrary.wiley.com/doi/10.1111/1468-2389.00169. doi:10.1111/1468-2389.00169.

[12] D. M. Truxillo, T. N. Bauer, R. J. Sanchez, Multiple Dimensions of Procedural Justice: Longitudinal Effects on Selection System Fairness and Test-Taking Self-Efficacy, International Journal of Selection and Assessment 9 (2001) 336–349. URL: https://onlinelibrary.wiley.com/doi/10.1111/1468-2389.00185. doi:10.1111/1468-2389.00185.

[13] K. Van Den Bos, R. Vermunt, H. A. M. Wilke, Procedural and distributive justice: What is fair depends more on what comes first than on what comes next., Journal of Personality and Social Psychology 72 (1997) 95–104. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.72.1.95. doi:10.1037/0022-3514.72.1.95.

[14] D. M. Truxillo, D. D. Steiner, S. W. Gilliland, The Importance of Organizational Justice in Personnel Selection: Defining When Selection Fairness Really Matters, International Journal of Selection and Assessment 12 (2004) 39–53. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.0965-075X.2004.00262.x. doi:10.1111/j.0965-075X.2004.00262.x.

[15] U. Konradt, T. Warszta, T. Ellwart, Fairness Perceptions in Web-based Selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply: Fairness in Web-based Selection, International Journal of Selection and Assessment 21 (2013) 155–169. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12026. doi:10.1111/ijsa.12026.

[16] A. Furnham, T. Chamorro-Premuzic, Consensual Beliefs about the Fairness and Accuracy of Selection Methods at University: Fairness and Accuracy of Selection Methods at University, International Journal of Selection and Assessment 18 (2010) 417–424. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2389.2010.00523.x. doi:10.1111/j.1468-2389.2010.00523.x.

[17] A. Mirowska, L. Mesnet, Preferring the devil you know: Potential applicant reactions to artificial intelligence evaluation of interviews, Human Resource Management Journal 32 (2022) 364–383. URL: https://onlinelibrary.wiley.com/doi/10.1111/1748-8583.12393. doi:10.

1111/1748-8583.12393.

[18] Q. M. Roberson (Ed.), The Oxford handbook of diversity and work, Oxford library of psychology, Oxford University Press, New York, 2013.

[19] T. E. Landon, R. D. Arvey, Ratings of Test Fairness by Human Resource Professionals, International Journal of Selection and Assessment 15 (2007) 185–196. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2389.2007.00380.x. doi:10.1111/j.1468-2389.2007.00380.x.

[20] G. S. Alder, J. Gilbert, Achieving Ethics and Fairness in Hiring: Going Beyond the Law, Journal of Business Ethics 68 (2006) 449–464. URL: http://link.springer.com/10.1007/s10551-006-9039-z. doi:10.1007/s10551-006-9039-z.

[21] S. Koivunen, T. Olsson, E. Olshannikova, A. Lindberg, Understanding Decision-Making in Recruitment: Opportunities and Challenges for Information Technology, Proceedings of the ACM on Human-Computer Interaction 3 (2019) 1–22. URL: https://dl.acm.org/doi/10.1145/3361123. doi:10.1145/3361123.

[22] H. Weerts, R. Xenidis, F. Tarissan, H. P. Olsen, M. Pechenizkiy, Algorithmic unfairness through the lens of eu non-discrimination law: Or why the law is not a decision tree, arXiv preprint arXiv:2305.13938 (2023).

[23] P. Sarkar, C. C. Liem, " it's the most fair thing to do but it doesn't make any sense": Perceptions of mathematical fairness notions by hiring professionals, Proceedings of the ACM on Human-Computer Interaction 8 (2024) 1–35.

[24] M. Wick, J.-B. Tristan, et al., Unlocking fairness: a trade-off revisited, Advances in neural information processing systems 32 (2019).

[25] A. K. Veldanda, F. Grob, S. Thakur, H. Pearce, B. Tan, R. Karri, S. Garg, Investigating hiring bias in large language models, in: R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Large Foundation Models, 2023.

[26] G. Malgieri, F. A. Pasquale, From transparency to justification: Toward ex ante accountability for ai, Brooklyn Law School, Legal Studies Paper (2022).

[27] K. Crawford, Can an algorithm be agonistic? ten scenes from life in calculated publics, Science, Technology, & Human Values 41 (2016) 77–92.

[28] M. Hildebrandt, Privacy as protection of the incomputable self: From agnostic to agonistic machine learning, Theoretical Inquiries in Law 20 (2019) 83–121.

[29] J. Simson, F. Pfisterer, C. Kern, Everything, everywhere all in one evaluation: Using multiverse analysis to evaluate the influence of model design decisions on algorithmic fairness, arXiv preprint arXiv:2308.16681 (2023).

[30] C. He, Y. Deng, A. Fabris, B. Li, A. Biega, Developing a fair online recruitment framework based on job-seekers' fairness concerns, arXiv preprint arXiv:2501.14110 (2025).