Fairness Beyond the Algorithmic Frame: Actionable Recommendations for an Intersectional Approach

Steven Vethman
Sciences Po Law School
Paris, France
TNO
The Hague, Netherlands
steven.vethman@sciencespo.fr

Quirine T. S. Smit
TNO
The Hague, Netherlands
quirine.smit@tno.nl

Nina M. van Liebergen
TNO
The Hague, Netherlands
nina.vanliebergen@tno.nl

Cor J. Veenman
Leiden University
LIACS
Leiden, Netherlands
TNO
The Hague, Netherlands
c.j.veenman@liacs.leidenuniv.nl

Abstract

Achieving fair use of AI systems is a multi-faceted challenge. Intersectionality, rooted in Black Feminist movements, is increasingly used to address the interconnected nature of discrimination such as racism, ableism, and sexism. Yet in AI research, intersectionality is often reduced to a narrow technical lens, focused on algorithmic bias between subgroups defined by protected attributes and addressed through fairness metrics. This algorithmic frame sidelines key aspects of intersectionality, such as power relations, social justice, and structural inequality. Still, AI experts play a central role in development and deployment, and therefore should act to limit unjust outcomes. This study offers actionable guidance for AI experts, grounded in a broader intersectional perspective. Through a thematic analysis of AI fairness papers on key aspects of intersectionality, evaluated through community engagement, we identify five themes with concrete recommendations: 1) insisting on collaboration in interdisciplinary teams, 2) embedding reflection and recognizing positionality, 3) approaching communities and facilitating co-ownership, 4) engaging with power dynamics and social context, and 5) assessing the framing and nuance of data and metrics. Participating experts noted barriers such as tech-optimism and fear of insufficient knowledge. Still, they valued the recommendations for communicating the importance of intersectionality and initiating more just AI practices. We call on AI experts to meet this challenge through interdisciplinary collaboration with diverse communities.

CCS Concepts

• **Social and professional topics → Computing / technology policy**; • **Computing methodologies → Artificial intelligence**.

FAccT '25, Athens, Greece

ACM ISBN 979-8-4007-1482-5/25/06
https://doi.org/10.1145/3715275.3732210

Keywords

Intersectionality, Discrimination, AI, Education, Survey, Co-design, Thematic Analysis, Recommendation, Collaboration, Participation, Position, Power, Measurement

ACM Reference Format:
Steven Vethman, Quirine T. S. Smit, Nina M. van Liebergen, and Cor J. Veenman. 2025. Fairness Beyond the Algorithmic Frame: Actionable Recommendations for an Intersectional Approach. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3715275.3732210

1 Introduction

The Dutch childcare benefits scandal, exposed in 2019, starkly illustrates how biased use of algorithmic systems harms marginalized groups. While unfairly targeting 26,000 families for fraud, it disproportionately affected already marginalized communities, particularly those with dual nationality, causing financial hardship and social stigma [51, 52, 62]. Such a contemporary algorithmic bias compounds the discrimination of marginalized communities, such as Black single mothers of Surinamese descent, who have historically been disregarded by the Dutch welfare system[1] [104]. A similar case emerged in France, where an automated fraud detection system used in welfare allocation disproportionately flagged individuals from low-income and immigrant backgrounds, leading to exclusion from vital benefits [4, 67, 82, 83]. These cases demonstrate how Artificial Intelligence (AI), or broader automated decision-making systems, can reinforce and propagate racism, sexism, and social inequality [3, 51].

AI fairness research recognizes the risks and harms of AI and aims to limit its discriminatory and unjust effects. Most work published on AI fairness focuses on group fairness [50, 76]: the aim to

[1] We capitalize "Black" to reflect its use as a socially constructed and political identity, rather than as a descriptor of skin color. In the Dutch context, Zwart (Black) carries similar anti-racist significance, though individuals of Surinamese descent may also identify as Bruin (Brown), depending on context [103].

minimize negative outcomes across demographic groups. Increasingly, this includes conceptualizations of fairness across subgroups who face multiple or intersecting forms of discrimination such as racism, sexism and ableism. These conceptualizations are often defined as intersectional (subgroup) fairness or intersectional bias in this community [44, 56, 59], in reference to the concept of *intersectionality* rooted in Black Feminist theory and practice [13, 28, 31].

This interpretation of *intersectional subgroup fairness* is criticized for being narrow and primarily technical [59, 73] as it mainly leads to algorithmic solutions, such as defining and optimizing for multidimensional fairness metrics. This focal point of AI fairness on the algorithmic frame or data frame (inputs, outputs and the model of the AI) distracts from more prominent issues of AI systems with respect to social justice that happen within the socio-technical frame (including humans and institutions surrounding the AI and their decisions) [55, 89]. In particular, the broader context in AI development and use is overlooked including power relations and the social context, which is central to both intersectionality and limiting the discriminatory and unjust effects of AI [28, 37].

The seminal work of Buolamwini and Gebru [26] further illustrates why a narrow interpretation of intersectionality falls short. The authors found that the AI-based image recognition systems made disproportionately more errors for women with darker skin. Many technical oriented papers cite this study in relation to their intersectional bias measurement and consequently continue to propose a bias mitigation techniques aimed at addressing intersectionality within the algorithmic frame by minimizing unequal errors [44, 71]. However, [26] started by understanding the social context of the problem, emphasizing that transparency and accountability reach beyond technical reports, and critically examined use of their measurement (also in [79]). A broader interpretation of intersectionality would advocate for analysing the social context of the data underlying the image recognition models. This analysis reveals unethical data labelling practices and a lack of diversity that have resulted in a process that particularly dehumanizes Black women [59, 68]. Moreover, by considering harms beyond the algorithmic frame in the social context where it is used—such as in mass surveillance and predictive policing—the disproportionate effect on dark-skinned women (and their lack of recourse) is incorporated [42]. On top of that, would a racially profiled Black woman care about more equal errors if she has limited opportunity to contest the errors that occur?

That is where the missed potential of intersectionality lies: the concept's core is about advancing social justice through examining and formulating actions based on the broader social context, inequalities, power relations, reflection, and hearing what marginalized voices prioritize [28]. In contrast to the narrower interpretation of intersectionality, which focuses on a descriptive understanding of discrimination related to intersecting identities, intersectionality is also understood as an approach, a framework for critical inquiry and practice towards social justice [109].

This interpretation of the broader *intersectional approach* for fairer AI views intersectionality as a lens for AI design, one that incorporates diverse perspectives and complex social contexts with the aim of promoting social justice. To help clarify the distinctions between interpretations of intersectionality, see a condensed summary in Table 1

This intersectional approach requires moving beyond the algorithmic frame, which is not the core of the AI expert's background, including their formal education and professional experience. Nevertheless, given their key position in the development and use of AI, they have a role and responsibility in the critical examination of how the AI systems they help create affect individuals and societal values [19, 79, 80].

In this study, we aim to support AI experts in fulfilling this responsibility for social justice by demonstrating how the broader intersectional approach can be integrated in their AI practices. This leads to the following research question:

- *What are the key actionable recommendations for AI experts to embrace the intersectional approach?*

We gather the key recommendations through a thematic analysis [21] on a tailored survey of AI fairness papers that incorporate key elements of the intersectional approach. Since the primary focus of AI fairness papers centres around the development and research of machine learning-based AI applications, we adopt this same scope for our study. For illustration, we reference the non-exhaustive list of high-risk AI applications outlined in the AI Act [101], which highlights relevant high-stakes AI projects. The choice of emphasizing *actionable* recommendations impacted our research design in two ways. One, we aim to include multiple examples and references per recommendation. This is to strike a balance between being general enough for operationalisation in multiple settings whilst also practical enough to inspire readers towards action. Two, we align our recommendations during the thematic analysis through inviting input from our target audience of AI researchers, data scientists and AI developers, whom we will hereafter refer to as AI experts.

In the next section, we present related work to contextualize intersectionality and its role within the AI fairness community. The third presents our methodology for collecting the recommendations and their community-based evaluation. The fourth presents the collected recommendations in five overarching themes. The fifth concerns the main insights from the evaluation by AI experts. The sixth presents a discussion including the limitations. Finally, the seventh concludes the paper.

2 Related Work

This section situates our contribution in related work. We commence with an introduction to the Black Feminist origin of intersectionality to gain further understanding of the concept as an approach. Thereafter, we discuss our contribution relative to adoption and critical examination of intersectionality in the AI Fairness literature.

2.1 Origin of the Intersectional Approach

Kimberlé Crenshaw coined the term "intersectionality" as she advocated that legal protection for discrimination should also be afforded for unique experiences from intersecting identities, in her case Black Women experiencing systemic racism and sexism [31, 32]. The concept is mainly associated to the Black Feminist Movement since Crenshaw brought it into public attention, yet it has been brought

Table 1: Comparison of Terminology, Frames, and Goals

	As used in [59]	Relating to frames in [89]	Goal as described in [73]
Intersectional (sub)group fairness	Narrow, Weak	Algorithmic frame	Optimizing for multidimensional fairness metrics
Intersectional approach	Broad, Strong	Socio-technical frame	Taking action towards social justice

forward by multiple movements as well as thinkers and activists before her [13, 27, 29, 86, 95]. A recent academic contribution that stands out is that of Patricia Hill Collins and Sirma Bilge [28]. Their work acknowledges the many definitions and interpretations of intersectionality and puts forward the *key elements* of the intersectional approach. One key element is that intersectionality is more than an analysis of discrimination based on intersecting identities; embedded within the concept is the objective of taking action towards social justice, which is often referred to as *critical inquiry and praxis*, encompassing both examination and action. We refer here to Iris Marion Young's definition of social justice, which emphasizes our shared responsibility for and the structural nature of (in)justice. According to Young, injustice occurs "when social processes put large groups of persons under systematic threat of domination or deprivation of the means to develop and exercise their capacities" [110, p.52]. Other key elements are investigating social context, positioning the work in social inequality and power relations, questioning one's own relation to the work, context and power relations, and being open to the complexity of social justice which ranges from intersecting identities, to hearing diverse voices and recognizing multiple kinds of knowledge. Informed by their work, these are the key elements we refer to in our research.

2.2 Intersectionality and AI Fairness Research

The bulk of papers in the AI fairness literature take a narrow interpretation of intersectionality as Kong [59] and Ovalle et al. [73] have established. However, there are recent AI fairness papers that incorporate key elements of the intersectional approach, such as power relations or the inclusion of diverse voices of the community participation, conceptually [15, 35, 39], applied in a use case [81, 87, 96] or even in their approach to data science education [8, 66, 80]. Most of the related work will be presented through the recommendations based on the literature survey. Below, we discuss two studies in particular that were foundational for our research and clarify how our contribution differs.

Kong [59] identifies the narrow interpretation of intersectional fairness in the AI fairness literature and consequently argues for three problems with this interpretation. She states that the field's focus on parity between statistical measurements of intersectional subgroups based on protected attributes is not enough because (1) the focus on attributes of race and gender diverts attention from the real problem being racism and sexism, (2) the focus on attributes rather than oppression creates a problem of arbitrary selection of intersectional subgroups, and (3) this view fails to address non-distributive aspects of fairness. Therefore, she advocates that this "weak" approach to fairness should be augmented into a "strong" fairness, one that acknowledges the structural nature of unfairness and aligns closely with a broader interpretation of intersectionality. Kong [59]'s contribution centres on illustrating the limitations of narrow interpretations of intersectionality in AI fairness research, whereas our work builds on this by developing actionable recommendations—evaluated and co-created with AI experts—to support those aiming to move beyond such limitations.

Ovalle et al. [73] performs a critical review of how intersectionality is discussed in 30 papers from the AI fairness literature. Their deductive and inductive analysis on the literature concurs with [59] that most papers reduce intersectionality to optimisation for fairness metrics and fail to discuss the social context and power. Their work establishes what was missing in the AI fairness literature related to key elements of the intersectional approach [28], and propose recommendations to researchers based on their identified gaps. Our work, building on their insights, focuses on how key elements of the intersectional approach were incorporated in the AI fairness literature and engage with AI experts how we can transform collected insights to actionable recommendations for them. Their work establishes what is missing in AI fairness research with respect to key elements of an intersectional approach [28], offering critical reflection and identifying key gaps. While they briefly propose directions for future research for each gap, our contribution builds on these insights by examining how such elements are already being engaged in the literature. We therefore shift the focus to surfacing and translating these practices into actionable recommendations, informed by input from AI experts.

Thus, the originality of our contribution lies in the fact that we guide AI experts beyond the critical review of their practices. Rather than focusing on their shortcomings or misinterpretations, we formulate insights from the literature into actions that they can and should take from their decisive role in AI development and research and align these with their input through community engagement.

3 Approach

In this section, we describe our approach to form actionable recommendations from the AI fairness literature, inviting AI experts to adopt an intersectional approach. To move beyond the algorithmic frame and make our recommendations actionable, we structure recommendations from interdisciplinary AI ethics conferences and align them with input from AI experts.

3.1 Recommendation Collection

Our approach to developing a set of actionable recommendations is based on analysing the recommendations found in interdisciplinary research on intersectionality. In order to structure the collected recommendations, we follow the well-established method for thematic analysis for qualitative research by [21, 22]. Below, we first explain the selection criteria for relevant papers, and then, elaborate on the phases of the thematic analysis method. In one of the phases, we use community engagement to align recommendations to AI experts' input, of which the description concludes this section.

3.1.1 Paper selection. The scope of our survey concerned two prominent multidisciplinary machine learning focused conferences: ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) and the AAAI/ACM Conference on AI, Ethics, and Society (AIES). We took all available papers at March 2024 of AIES (2018-2023) and FAccT (2019-2023) that mentioned intersectionality in their full text. Altogether, this gave a total of 268 papers.

For those papers, we divided the papers among the authors and read the titles and abstracts and looked at whether they fit the broader interpretation of the intersectional approach. This resulted in the inclusion of papers that refer to key elements of the intersectional approach as defined by Collins and Bilge [28] (see 2.1), where any papers in doubt were read and discussed through using the guiding questions of [73, p. 59] in Table 3. In case of disagreement, the paper was included. This resulted in 63 papers.

3.1.2 Thematic Analysis. To extract the key actionable recommendations from the selected papers, we chose the methodology of a thematic analysis (TA), which is used to identify, analyse and report on the patterns or themes in written text or speech [21]. This is done through *coding* (labelling through interpretation) of the *data* (for us, text snippets in the selected papers that describe proposed actions or explicit recommendations), and, in turn, through iterative analysis of patterns in these codings arriving at *themes* (the overarching key recommendations).

To fit our research needs, we selected the inductive reflexive TA known from Braun and Clarke [21], over two other common variants of TA: coding reliability TA and codebook TA [23]. We highlight two key features to ground this choice. One, the inductive reflexive TA starts with letting the data speak for themselves without categorization from previous frameworks. This fits our aim to let the authors of the selected papers have a voice in the patterns to be established. Two, reflexive TA regards the subjective interpretation in coding (by the authors of this work) as valuable processing that builds on previous experiences and backgrounds rather than a bias to be avoided. This deviates strongly from the desired homogeneity and inter-annotator agreement in coding reliability TA. To provide structure to the flexible method, Braun and Clarke [21] defined a process of six phases. Following this process, we started by familiarising ourselves with the dataset by reading and iteratively decided on a coding strategy (1). Then, we generated initial codes by interpreting and paraphrasing the core of explicit recommendations for actions written in selected papers (2). We searched for themes by grouping the codes into clusters, which we iteratively refined through multiple collective discussions. (3). After, we reviewed, refined and named the initial themes to arrive at five over-arching themes and formulated sub-themes as actionable recommendations (4). We evaluated both with the help of AI experts (5), see in section 3.2. Finally, we aligned our recommendations based on their input, and through the act of writing finalized them in this paper (6). This way, we structured 206 extracted recommendations into five themes. More details and intermediate results are described in Appendix B.

3.2 Evaluation Session

To enhance the actionability of our recommendations for AI experts, we incorporated community engagement within our thematic analysis. For this, we organized an interactive workshop and an open-ended survey.

The interactive workshop was hosted at a Dutch research institute. We had 22 participants on-site and five participants online. All were informed about the study and gave informed consent to contribute anonymously. We have chosen to not collect personally identifiable data, yet describe hereunder the perceived or known representation of the group. The participants disciplines ranged from technical, social to managerial, with a large variety in roles such as human rights lawyer, project manager, data scientist, research group manager, intern, governance expert, AI researcher, junior consultant to senior researcher. Although this set of participants had a broader range of titles/roles than we defined as our direct target audience in section 1, all of them have played a part in interdisciplinary AI development or research teams. We valued this additional variety in disciplines, which fit our own recommendation in section 4.1 such that different interpretations, language and roles could be discussed. Dutch nationals formed the majority of participants, along with at least five international participants. Concerning gender diversity, we noticed that over half of the people used she/her pronouns. To safeguard privacy, we do not disclose any perceived or disclosed representation on unobservable or private memberships such as being queer or neurodiverse or having disabilities. Concerning age diversity, participants ranged broadly from their twenties to their sixties, with a skew toward younger individuals.

The workshop was organized in three parts. First, an interactive introduction: defining intersectionality together, watching an educational video on intersectionality and finally discussing an example of image recognition with the narrow and broader interpretation of intersectionality [26, 79]. Second, the participants were split into two groups on-site and one group online. Each group, hosted by one of the authors to facilitate and take notes, discussed two different themes of the recommendations. Starting with a discussion on the overarching recommendation (theme) and the link to intersectionality, each group finished with reviewing three possible actions related to the recommendation. The three possible actions refers to intermediate versions of the actionable recommendations written for each theme in section 5. Third, a paper survey was presented to ask the participants on four aspects of intersectionality to acquire feedback on the actionability of the recommendations.

Q1. How important do you find intersectionality?
Q2. To what extent do you find the social science theory of intersectionality understandable?
Q3. In previous discussions with data scientists on intersectionality, some felt criticized by the content. We're curious to hear your thoughts: do you feel similarly? To what extent do you feel attacked or have the sense that you're not doing well?
Q4. To what extent are the recommendations practical enough?

Discussing the recommendations with AI experts influenced the authors of this paper on the tone, formulation and framing of the actionable recommendations. Insights from the session are therefore interwoven in our actionable recommendations, whilst two main insights are discussed in section 5.

4 Recommendations

In this section, we present our recommendations based on our thematic analysis of the literature. The collected recommendations are offered in five overarching themes. Each theme is first briefly presented as its key message and an elaboration on its link to intersectionality. Thereafter, actionable recommendations are described to facilitate AI experts to take the first step in incorporating the intersectional approach.

Figure 1 represents a schematic overview of the recommendations. Firstly, as AI experts are centred in AI development and practice, they have the decisive role to insist on the interdisciplinary collaboration that AI fairness requires. Secondly, these interdisciplinary teams should discuss and document their position in society, to be aware of the perspectives they bring to their practice. Thirdly, invite people at risk of AI harm to voice priorities and concerns and propose co-ownership of their participation process to them. Fourthly, together with affected communities, the interdisciplinary teams should analyse the power relations between those using the AI, benefiting from the AI and those potentially harmed by the AI, as well as their context of systemic societal inequalities. Finally, given all these perspectives and insights, discuss how and if the opportunities and limitations of measurement and technological solutions with data and metrics align with the goal of social justice

See Appendix A for a table that provides an easy overview of all actionable recommendations per theme.

4.1 Collaboration and Role

As AI experts are centred in AI development and practice, they have the decisive role to insist on the interdisciplinary collaboration that AI fairness requires.

4.1.1 Importance to intersectionality. An intersectional approach requires rich dialogue and multiple perspectives. With the help of other experts, it is easier to look at fairness beyond the algorithmic frame. Recall that the intersectional lens particularly incorporates the nuance of power hierarchies. AI experts should recognize that they are given the responsibility (power) of AI fairness most of the time and use this role to assert the need for interdisciplinary teams.

4.1.2 Collaborate with multiple disciplines before going into technical details. To do justice to AI fairness, a variety of expertise (think of social science, AI governance, and domain expertise) are needed to understand what the problem and/or related goal are before one defines them in terms of data and technical implementations [61, 92]. Collaborating with different disciplines takes time to understand each other's language. Organize a session to present each other key concepts from each discipline, share different interpretations and find common terminology for the key concepts [66]. Similarly, the choice of research methodology also require an open discussion. The perceived value of quantitative and qualitative research methods or testing of AI differs per discipline, where the education of AI and data-science focus on quantitative methods [8]. Games, prototyping or designing visual diagrams of different scenarios of AI impacts are different ways to establish a common language and set the scene for a creative joyful collaboration [10, 19]. The applied nature of these exercises makes assumptions, values and expectations tangible, visible and engaging [8]. We tend to perceive potential AI harms first from our own experience, which makes storytelling a useful tool to include more perspectives [66].

4.1.3 Use central role of AI experts to invite other disciplines and share responsibilities. Data scientists and AI researchers are often given the responsibility to handle AI fairness in practice, so they have the leverage to make a stand on development, research or practice with integrity [80, 91]. Stand your ground that considerations around fair use of AI requires interdisciplinary collaboration throughout the lifecycle and argue to managers—or those with power on deciding the team composition— the need to invite colleagues with different areas of expertise [2, 20]. Boag et al. [18], who examined the relationship between "AI Ethics" and employee activism, demonstrate further possibilities. They demonstrate how AI experts can organize collectively to exert influence and provide concrete strategies for such collective action, including examples like organizing strikes.

4.1.4 Dedicate time and effort to create a psychologically safe environment. Conversations between different disciplines are bound to start with misunderstandings and disagreement before common language and shared goals are established. Therefore, set out to create an open, informal, and trusted setting [8, 57]. It may be advised to avoid setting a specific goal during the first discussions where interdisciplinary teams come together; a goal such as *the problem definition needs to be done by the end of this meeting* already sets the frame that critical voices that disagree with the majority or raise ethical voices are not appreciated [20, 57]. Where collective decisions are required, consider deliberative approaches alongside voting mechanisms, in order to give voice to minority priorities [19]. Establish and document common values, expectations and norm for communication styles [19, 41]. If disagreement on priorities is inevitable, allow those whose wishes are not met to air their hesitations and document these [19].

4.2 Position and Reflection

Interdisciplinary teams should discuss and document their position in society and reflect which perspectives are heard and which are still left unheard.

4.2.1 Importance to intersectionality. By taking your position and reflection seriously, you dare to be transparent on what you see as priorities based on your background and experience, whilst realizing that unique diverse experiences exist. State your doubts on whose voices might be missing in the team and how you aimed to still reach those voices (but perhaps failed). This positioning may at first glance appear exposed or self-doubting relative to common expressions of "bias-free" technology. Yet, this reflective vulnerability also demonstrates a strong open stance of a willingness to learn and accountability, which invites grace when any issues are detected.

4.2.2 Write a positionality statement and reflect on it. Dedicate time at the start of the project to sit down with the whole team and reflect

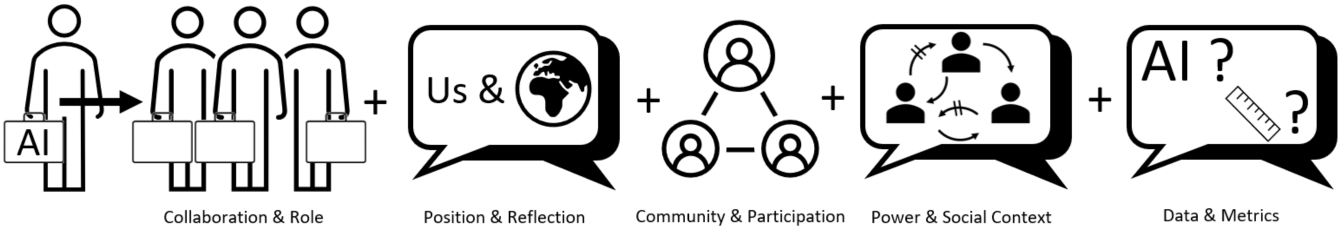

Figure 1: Overview of recommendations on how to do AI fairness with an intersectional approach, from the responsible role of the AI expert, the building of a multi-disciplinary team, the reflection on the team's position in society, the participation and co-ownership of relevant communities, to the consideration of power relations and the social context, and the role of data and metrics.

on the team's and AI system's position in society [48, 57, 73, 89]. To determine which positions are relevant in your setting, consult with domain experts familiar with the context in which the AI system is situated. What discrimination or equality concerns have previously been raised in this domain or specific setting? Discuss which roles are present around the AI system and commenting on the privilege and power relations you perceive (e.g. sponsor, the person with ultimate responsibility, designer, user, affected people or parties) [57]. An intersectional question would be: if we have for example women or queer representation in the team or the advisory panel, do we have only the most privileged subgroup of the marginalized group? Writing this statement may also frame your contribution, as you can reflect upon: based on our representation, what is our role to play? See also our own positionality statement for an example of this.

4.2.3 Document perspectives and decisions throughout the lifecycle of AI. Every time your team needs to make a decision, reflect on your positionality statement, and document how this may affect the decision [48, 73]. Write down the varying perspectives and opinions in the team on each possible alternative or choice as well as the final decision made. This form of transparency fosters open communication and shared accountability. Not all opinions and expectations need to be met or agreed before decisions are made, yet all these voices need to be heard and documented such that any systemic patterns can also be examined [19]. Important decisions with respect to intersectionality are concerning the position of the AI system and social inequalities, the intended use and the choice and definition of social categories to include. More on these aspects follows in sections 4.4 and 4.5.

4.3 Community and Participation

Invite people at risk of AI harm to voice priorities and concerns and propose co-ownership in the participation process.

4.3.1 Importance to intersectionality. The intersectional approach acknowledges the variety of voices and that some are heard more than others. Therefore, it is crucial to invite communities at stake to participate meaningfully throughout. That is, not only in the last alignment, but from the start in setting the goal of the AI project.

4.3.2 Invite communities to co-own the participation process. The paramount recommendation for community participation was to give the community a meaningful voice and control in the participation and full transparency on this matter [19, 36, 48, 55, 61, 65, 81, 87]. Clear communication and agreements to what extent communities share or have leadership or co-ownership of outcomes and decisions is crucial to establish informed consent for participation [36]. This cannot be an afterthought as meaningful participation takes time and space [7, 55, 81]. This investment in time and effort starts with process of identifying and reaching out to impacted communities and/or their representatives for participatory development [19]. [19] discusses multiple approaches to reach communities. Participatory AI projects should begin with AI literacy efforts, as participants need a foundational understanding of what AI is, and what it is not, in order to meaningfully engage in discussions about how such systems may affect their lives [99]. The co-designing exercises suggested for interdisciplinary collaborations are tangible suggestions for community participation as well [19, 38]. Flexibility to change vital parts of the AI project is required for community participation to become worthwhile. [55] show clear lessons learned based on an applied use case, where global and generic aspirations of the project needed to make space for first the impact and effectiveness at a local scale. Other examples of co-design can be found in [74] concerning AI and ableism and in [24] on AI speech technologies, racism and ageism.

4.3.3 Make participation financially sustainable for communities. As we value community experience and knowledge they contribute, make sure to compensate them accordingly [48, 61]. Participation needs to be mutually beneficial and financially sustainable on the long run [36]. To ensure a fair compensation rate, an external party can be involved to determine the appropriate amount [74].

4.3.4 Design a mechanism where impacted communities can safely voice concerns. As is generally known, making AI systems 100% bias free is not possible. Any AI harm that may slip through your mitigation strategies, therefore requires a pathway to be heard. Effective participatory design requires careful attention to the social context and power relation such that impacted communities can safely voice concerns. [81] show how sometimes it is crucial to allow communities to voice concerns collectively and anonymously, as users or impacted people may be ashamed or otherwise disadvantaged to hold systems accountable individually on their own [81]. Voicing concerns does not need to be a passive feedback mechanism, where one waits for problems to arise. Start from defining harms together with communities to see what could be measured and monitored (rather than starting from definitions of statistical fairness) and facilitate co-designing sessions where communities can make concrete what it means when error may occur [6, 73, 81, 89]. In other words, take diligent care to prevent harms upfront, but give space and power to voiced errors that slip through.

4.4 Power and Social Context

Interdisciplinary teams, together with communities, should analyse the power relations between those creating, researching, using, benefiting from and those (potentially) harmed by the AI, within their social context.

4.4.1 Importance to intersectionality. Positioning the AI system within power and social context with diverse voices (disciplines and communities) is key to adhering to the goal of social justice, which is central to intersectionality. Centring marginalized voices and acknowledging the systemic nature of discrimination helps to prevent regarding "fairness" as a single isolated add-on or an afterthought in your AI project.

4.4.2 Position the AI within social context and define the present power relations. Societal implications of AI projects are better addressed when they are positioned in their historical, social and cultural context and power relations [2, 8, 15, 57, 79]. Consider the famous COMPAS case, where [5] showed through investigating the error distribution that AI used in the justice system propagated systemic racism. There, discussing power relations and social context within the U.S. justice system may shift priorities away from improving the AI system. Within the context of the industrial prison complex, the power dynamics present in prison, the justice system and historical racial profiling, marginalized communities and their advocates have voiced priorities such as less prisons and more contestability [33]. Therefore, come together with the interdisciplinary team, including communities or community representatives, and discuss who took the initiative on the project, who will benefit, who may be harmed and who has a voice [54]. Impact assessment such as FRAIA [46], HUNDERIA [30] and FRIA [64] are tested and anticipated opportunities if a more extensive and structured format to discuss potential benefits and harms of AI systems fits the project. We refer back to communication suggestions from [8, 57] as discussions on power and privilege are conversations where disagreement and critical opinions should be handled with care. Identify groups who are relevant but absent or under-represented in the process, and reflect on whether their exclusion is shaped by social inequalities or historical context .[57] demonstrate that through prototyping your AI system or research and writing an accompanying narrative that discusses who is involved and what their roles are can help reflect which voices are still missing.

4.4.3 Redefine concepts with power and social context in mind. With an understanding of the power relations and social context, it is crucial that the team remains open for priorities or goals of the AI research or the AI system to change [8, 9, 55]. Priorities are often centred around key concepts such as accountability, responsibility and fairness. Take time with the team and the community to redefine them with power and social context in mind (inter alia [8, 73]). It is advised to revisit these interpretations iteratively due to the dynamic nature of these concepts as well as shifting power relations and social context [38]. We highlight here Klumbytė et al. [57], who share multiple approaches (intersectional feminist methodologies) that are specifically designed for interactive interdisciplinary discussions on the design of machine learning within a context of power. [48, 57, 61, 81] are examples where critical analyses of power and social context have influenced their AI fairness research. [57, 81] redefined the concept of accountability, by analysing who has the ability to respond. In [81], they show how a technical solution for accountability did not meet the social-political reality of people taking loans in India, and propose a collective yet anonymous form of accountability for effective redress. Moreover,[12, 48] demonstrate how creating a common understanding in your team of race being a social construct, supports avoiding common mistakes such as misattributing racial categories as causal mechanisms, reifying race as a natural category and misidentifying race rather than racial stratification as the root cause of disparities. See [1] for an applied example concerning the social context of gender and sex.

4.5 Data and Metrics

Given all these perspectives and insights, discuss how and if the opportunities and limitations of measurement and technological solutions with data and metrics align with the goal of social justice.

4.5.1 Importance to intersectionality. An intersectional approach acknowledges the political nature and power of data and metrics. Within the goal of social justice, reflect on the framing of your intended use with data and metrics. Dare to ask beyond how AI can help, and ask the zero question: is AI suitable for this problem at all? If so, demonstrate how your use of data and metrics has impact beyond the algorithmic frame. Show how you recognize the complexity of how systemic forms of discrimination may interact and are embedded throughout the AI system. Document clearly how you establish the added value and limitations of your data and metrics, for example, through qualitative methods or community participation.

4.5.2 Be critical on your objective with data and metrics. Throughout this paper we have invited AI experts to collaborate within an interdisciplinary team and with diverse communities. This last recommendation requests their shared goal to be beyond technically solvable issues and to be set towards achieving social justice [54, 55, 73]. Ask the zero question, examine *whether* rather than *how* AI system should be developed or implemented [55, 73]. Measurement with data and metrics can still have value for social justice, see for example the work of [26] where intersectional bias measurement had a signalling function to show how systemic racism and sexism is embedded in image data and image recognition tooling. Framing here is essential, as the same authors requested caution for their measurement functioning for certifying AI as bias-free [79]. [17] argue for similar caution when aiming to optimizing for seemingly neutral or objective benchmarks. Instead of bringing social justice into AI systems that have a secondary goal such as credit scoring, hiring or detecting fraud, AI can also be developed and (co-)created for the sole purpose of social justice. Through feminist participatory design Suresh et al. [96] develop an AI system to support activists monitoring feminicide. They highlight the challenges posed by incomplete and infrequently updated data of feminicide, which obscures the systemic nature of violence, and emphasize how the lack of data actively disempowers women. In response, they co-designed datasets and machine learning models to aid in the collection and analysis of feminicide data to support activist efforts. Their work shows how datasets can contribute to oppression and actively fights this by collecting inclusive more data.

4.5.3 Augment quantitative approaches with qualitative research and participatory design. Even within the right framing, a challenge to capture the complex societal phenomena of structural inequality within data and metrics remains. Keep in mind that forms of discrimination interact such that within one group there can be many different experiences [100]. Go beyond single axis thinking with your measurement, or check with a large variety of organizations representing different subgroups whether your AI-solution is also useful for them [96]. Tomasev et al. [100] also list specific considerations for complexity that arises when a group is marginalized for unobservable characteristics. Quantitative data and metrics are likely to capture only part of the phenomenon, and may also be undesirable due to privacy concerns surrounding sensitive data. Therefore, they cannot be relied upon alone [48, 74]. We recommend complementing them with qualitative approaches, such as interviews and focus groups, ideally involving communities or representatives of those communities [78, 96].

4.5.4 Document clearly on the intended use and limitations of data, model and metrics. During your efforts to capture the complexity of systemic inequalities in data and metrics, providing documentation on intended use is crucial. This includes providing thorough documentation on the data used, the researcher's goals in collecting the data and creating the model, and note potential users and stakeholders who could be negatively impacted by model errors or misuse [1, 10, 14]. Next to that, be transparent on your efforts for accountability by transparent communication on any side effects, which includes how they may affect vulnerable people as well as what you currently do to prevent them [6, 53]. Do not be negligent and dismiss them as unlikely or unintended consequences and stop there [19]. This recommendation for humility and transparency also holds for performing audits on AI. [79] demonstrate how important it is to not oversell the value or reliability of an audit to certify the ethical use of an AI.

5 Community Insights

In this section, we discuss the two main insights from the evaluation session with AI experts. The primary outcomes concern the work environment and the fear of not knowing enough. Note that we also reflect on these further in the discussion section hereafter.

5.1 Work Environment

Our recommendations start and have stressed the role of AI experts in bringing the intersectional approach to AI fairness practice. However, participants have voiced that they expect AI experts' influence to bring in critical examination of the goal of the project or proposing non-technical alternatives may be restricted by their work environment. One participant of the evaluation workshop shared: "I notice that in most projects during my career, we aim to do the most as possible with the data available, rather than questioning whether doing the analysis at all, will provide a sufficient and meaningful answer to the problem." Other participants noted that quantitive measures are often valued higher than qualitative methods, also by external stakeholders, as well as that sometimes an AI solution is necessary given the funding and is therefore a goal on its own.

On the other hand, participants also noted on two potential facets how the recommendations may be used to tackle obstacles from the tech environment. Firstly, the recommendations with its examples and communication strategies could aid in articulating the importance of community participation, social context and interdisciplinary collaboration, among others, to project stakeholders and funding decision-makers. Secondly, we learned through community participation that the recommendations were perceived to facilitate decision-making about AI that is more aligned with social context and stakeholders. In turn, this would make any pursued AI projects land better in society. They noted that funders of AI projects eventually care about these success stories; careful consideration on which AI projects to pursue may therefore be aligned with the goal.

5.2 Fear of Not Knowing Enough

Although our recommendations provide actions and steps for AI experts to follow, participants of the workshop noted that unfamiliar language, concepts and perspectives also caused a sense of unfamiliarity and unreadiness. Particularly, they voiced that a fear of not doing well and having blind spots makes them hesitant to apply the intersectional framework at all; they want to do it "perfectly" and in a structured, coordinated way. This was also expressed through many questions, such as: "So I know I am a white cishetero man, doing research on bias, what does that mean for how I do bias research?", "There can be so many intersections. How do you check all of them?" Additionally, the participants expressed a fear of finding blind spots during the process. Here they discussed a linear thinking approach: they are used to know beforehand which tasks need to be performed and in which order, which ideally is complemented with a checklist. However, the intersectionality framework also asks for adaptation during the process, think of the co-ownership in the community participation.

6 Discussion

In this section, we discuss reflections on the value of our contribution as well as its limitations.

From the start of this paper we have advocated that AI experts have a decisive role in bringing intersectional fairness concerns beyond the algorithmic frame to the AI harms voiced by marginalized communities. We recognize, however, as the participants of the evaluation session had voiced, that their influence is dependent on their work environment. Especially with the omnipresence of tech-solutionism or algorithmic idealism [35], we acknowledge the challenge that (although motivated) most researchers and practicioners are still subject to funding or higher management. We argue that our actionable recommendations allow AI experts to start small. Invite someone from another company, department or research group to bring another perspective, write a positionality statement with the team and reach out to a few civil society organizations that represent communities. AI fairness is a marathon, you cannot wait for the perfect conditions to start practice your running. To facilitate further alignment within the work environment of AI development, we foresee opportunity for future research to embed the actions in our recommendations in widespread iterative approaches for AI/software development such as Agile Scrum and CRISP-DM [88, 105]. AI experts can also take in our recommendations beyond the work environment of an institute or private organization. Rather than starting with AI experts, asking for interdisciplinary collaboration and community engagement, an intersectional approach to AI fairness is also very suitable to start at civil society. Such an operationalization of intersectionality is also what [35] strive for in their suggestion to redistribute AI power. They provide examples where community-led academic-activist collaborations use AI education, evaluation and design to address historical wrongdoings that affect current and future opportunity structures.

Moreover, we also recognize that our recommendations request AI experts to go beyond their comfort zone of their previous experiences and education. We invite AI experts to bear some of that discomfort as a level of unfamiliarity, as the stakes of AI fairness are per definition high, which is best heard through critical voices who have experienced AI harms. We stress that AI experts do not need to and should not go through these lessons alone, as diverse perspectives are essential. Nor do AI experts need to become experts on power and social context. Yet, as it is the red line through all our recommendations that embody the intersectional approach, we dare argue that AI experts do have the duty to invest in educating themselves to the level that they can collaborate within interdisciplinary teams and communities that bring their expertise on power and social context. As our recommendations put forward the importance of soft skills for AI fairness (due to the value of interdisciplinary collaboration, community participation, qualitative research methods), we see potential for critical data science education curricula (such as discussed in [8, 70]) to take in our concrete recommendations; bringing the comfort zone of new AI experts beyond the algorithmic frame.

To enhance the actionability for AI experts further, we also see opportunity for future research for creating a learning environment for critical discussions. Inspiration may be gathered from concepts such as psychological safety, pioneered by Edmondson [40, 41] who defines psychological safety as the team environment where members can be with candour to take risks, express ideas, speak up with questions and admit missteps. Another relevant practice already situated within the goal of social justice is that of calling in, currently propelled by Dr. Loretta J Ross Ross [84]. Her recent publication materializes her advocacy work (e.g. [85]) where calling in is championed as an approach to invite change through compassion rather than expecting that someone has already grown. Their approaches may aid in creating the environment necessary to call in AI fairness experts as allies to move beyond the algorithmic frame to social justice, whilst limiting the chance for polarizing based on past practice.

6.1 Limitations

Many perspectives were heard through the community engagement and the literature survey. However, there are other perspectives we did not hear. Firstly, the participants of the evaluation session were all AI experts, that although occupying a variety of roles and having different educational backgrounds, were European and worked at the same Dutch research institute. On top of that, due to the self-selective nature of the evaluation session, the participants were also all AI experts who value and are interested in learning more about intersectional fairness in AI. Hence, we should take note that our definition of actionable recommendations for AI experts is biased towards European (Dutch) and motivated AI experts. We argue, however, that the impact of the limitation of motivation is limited. It is also these motivated AI experts that are most likely to act upon our invitation, which through them may ripple further.

Secondly, the recommendations stem entirely from academic sources, and specifically from the selected conferences, which naturally results in an under-representation of relevant activist voices publishing outside academia. Some of the included studies are authored by scholars who also act as activists or represent civil society groups working directly with people harmed by algorithms, while others were explicitly co-created with such communities (e.g. [36, 73, 81, 96]). Still, our interpretation and translation of their insights into actionable recommendations have not yet been explicitly evaluated against the priorities of activist communities or those affected by AI. Future research should build on our work by including such an evaluation and/or by explicitly drawing from a broader range of disciplines beyond AIES and FAccT such as critical legal studies, political science, or social sciences like gender and race studies. This could bring the recommendations even closer to the core principles of an intersectional approach, which emphasize the value of diverse knowledge systems and the centring of marginalized voices. In turn, this may improve their applicability within the proposed interdisciplinary teams and ensure better alignment with the perspectives of those who have experienced harm.

Finally, throughout the paper we use terminology to describe groups of people. We have made a conscious effort to use the language that is preferred by the people of the communities we are referring to. We hope that our use of language serves as a vessel to get our point across. Whilst continuing to educate ourselves, we welcome hearing any unintended sense of exclusion we have caused due through word-choice.

7 Conclusion

Much of the AI fairness community currently engages with intersectionality through a narrow, technical lens that focuses on data, model and outputs, also known as the algorithmic frame. This results in efforts that often target algorithmic bias between subgroups defined by protected attributes such as sex, nationality, and skin colour. Yet intersectionality, grounded in Black Feminist thought, offers a much broader framework that can support AI experts in addressing the structural dimensions of unfairness through a social justice lens. Through a thematic analysis of a tailored literature survey and community engagement, we have formulated five actionable recommendations for AI experts. (1) As AI experts are centred in AI development and practice, they have the decisive role to insist on the interdisciplinary collaboration that AI fairness requires. (2) As the team takes on a responsibility towards social justice, it is key to position themselves within society and reflect which relevant voices are heard and unheard. (3) Through meaningful community participation, the people at risk are also invited to safely voice concerns, co-own the process of their participation and be financially compensated. (4) Then, the interdisciplinary teams, together with communities, should analyse the power relations between those creating, researching, using, benefiting from and those (potentially) harmed by the AI, within their social context. (5) Given all these perspectives and insights, discuss how and if the opportunities and limitations of measurement and technological solutions with data and metrics align with the goal of social justice. We invite AI experts willing to integrate the intersectional approach to embrace any discomfort experienced and hope that our actionable first steps help with this process. We are on that journey ourselves.

8 End Matter Sections

8.1 Positionality Statement

As we call for positionality in AI research, we also reflect on our own. We are AI researchers trained in mainly computational and quantitative sciences, currently working in the field of responsible AI. Our motivation to bring intersectionality into AI practice was shaped through critical engagement with foundational work in the field (e.g., [59, 73]) and group discussions. Although the bulk of our formal education lies outside the social sciences, some of us have actively taken relevant courses (such as on human rights and psychological safety), been part of employee resource groups, and learned informally through queer and activist communities. We also strive to collaborate with social science experts and have sought their feedback on this work. In forming our approach, we questioned whether we could do justice to the Black Feminist roots of intersectionality. While some of us have experienced intersecting forms of discrimination, we also acknowledge our many privileges. We are all based in the EU, have academic backgrounds, and are mostly in contact with others in similar contexts. These positionalities shape both our insights and our blind spots. This led us to refine our goal: to support AI experts, ourselves included, in meaningfully engaging with intersectionality. We take seriously our responsibility to contribute to social justice in AI, drawing on frameworks of critical self-reflection and "calling in" [84, 110]. We see this work as a step in an ongoing process and invite feedback, particularly from critical voices and civil society perspectives.

8.2 Author Contributions

The contributions of the authors were as follows, using the initials of first and last name.

- Conceptualization: SV, QS, CV
- Methodology: SV, QS
- Formal Analysis: SV, QS, NL, CV
- Evaluation Session: SV, QS, NL
- Writing – Original Draft: SV, QS, NL
- Writing – Review & Editing: SV, QS, NL, CV

8.3 Competing Interests

The authors have no known competing interests.

8.4 Acknowledgements

To start, we acknowledge that this work is part of the DIVERSIFAIR project that aims to deliver educational resources to foster inclusive AI practice through an intersectional approach. The DIVERSIFAIR project is a result of the Erasmus+ grant (101107969), and is therefore co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Educative and Culture Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Moreover, we thank the whole DIVERSIFAIR team for the constructive collaboration and in particular the reviews from Raphaële Xenidis and Ilina Georgieva. Next to that, we thank the anonymous FAccT reviewers for their valuable input. Finally, we thank all the artists and thinkers, among whom Teju Cole, Carrie Mae Weems, James Baldwin, and civil society organizations such as Bits of Freedom and Racism Technology Center who have and continue to inspire us. And Ivy.

References

[1] Kendra Albert and Maggie Delano. 2021. This Whole Thing Smacks of Gender: Algorithmic Exclusion in Bioimpedance-based Body Composition Analysis. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 342–352. doi:10.1145/3442188.3445898

[2] Andrea Aler Tubella, Dimitri Coelho Mollo, Adam Dahlgren Lindström, Hannah Devinney, Virginia Dignum, Petter Ericson, Anna Jonsson, Timotheus Kampik, Tom Lenaerts, Julian Alfredo Mendez, and Juan Carlos Nieves. 2023. ACROCPoLis: A Descriptive Framework for Making Sense of Fairness. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1014–1025. doi:10.1145/3593013.3594059

[3] AlgorithmWatch and Bertelsmann Stiftung. 2020. Automating Society Report 2020: Taking Stock of Automated Decision-Making in the EU. https://automatingsociety.algorithmwatch.org/ Accessed: 2025-01-10.

[4] Amnesty International. 2024. France: Discriminatory algorithm used by the social security agency must be stopped. https://www.amnesty.org/en/latest/news/2024/10/france-discriminatory-algorithm-used-by-the-social-security-agency-must-be-stopped/ Accessed: 2025-05-04.

[5] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing Accessed on 21 January 2024..

[6] Pınar Barlas, Kyriakos Kyriakou, Styliani Kleanthous, and Jahna Otterbacher. 2021. Person, Human, Neither: The Dehumanization Potential of Automated Image Tagging. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 357–367. doi:10.1145/3461702.3462567

[7] Julia Barnett and Nicholas Diakopoulos. 2022. Crowdsourcing Impacts: Exploring the Utility of Crowds for Anticipating Societal Impacts of Algorithmic Decision Making. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Oxford United Kingdom, 56–67. doi:10.1145/3514094.3534145

[8] Jo Bates, David Cameron, Alessandro Checco, Paul Clough, Frank Hopfgartner, Suvodeep Mazumdar, Laura Sbaffi, Peter Stordy, and Antonio De La Vega De León. 2020. Integrating FATE/critical data studies into data science curricula: where are we going and how do we get there?. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 425–435. doi:10.1145/3351095.3372832

[9] Bilel Benbouzid. 2023. Fairness in machine learning from the perspective of sociology of statistics: How machine learning is becoming scientific by turning its back on metrological realism. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 35–43. doi:10.1145/3593013.3593974

[10] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 610–623. doi:10.1145/3442188.3445922

[11] Sj Bennett, Caroline Claisse, Ewa Luger, and Abigail C. Durrant. 2023. Unpicking Epistemic Injustices in Digital Health: On the Implications of Designing Data-Driven Technologies for the Management of Long-Term Conditions. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 322–332. doi:10.1145/3600211.3604684

[12] Sebastian Benthall and Bruce D. Haynes. 2019. Racial categories in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, Atlanta GA USA, 289–298. doi:10.1145/3287560.3287575

[13] Michele Tracy Berger and Kathleen Guidroz (Eds.). 2009. *The Intersectional Approach: Transforming the Academy through Race, Class, and Gender*. University of North Carolina Press, Chapel Hill.

[14] A. Stevie Bergman, Lisa Anne Hendricks, Maribeth Rauh, Boxi Wu, William Agnew, Markus Kunesch, Isabella Duan, Iason Gabriel, and William Isaac. 2023. Representation in AI Evaluations. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 519–533. doi:10.1145/3593013.3594019

[15] Abeba Birhane, Elayne Ruane, Thomas Laurent, Matthew S. Brown, Johnathan Flowers, Anthony Ventresque, and Christopher L. Dancy. 2022. The Forgotten Margins of AI Ethics. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 948–958. doi:10.1145/3531146.3533157

[16] Emily Black, Hadi Elzayn, Alexandra Chouldechova, Jacob Goldin, and Daniel Ho. 2022. Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1479–1503. doi:10.1145/3531146.3533204

[17] Borhane Blili-Hamelin and Leif Hancox-Li. 2023. Making Intelligence: Ethical Values in IQ and ML Benchmarks. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 271–284. doi:10.1145/3593013.3593996

[18] William Boag, Harini Suresh, Bianca Lepe, and Catherine D'Ignazio. 2022. Tech Worker Organizing for Power and Accountability. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 452–463. doi:10.1145/3531146.3533111

[19] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 425–436. doi:10.1145/3461702.3462612

[20] Karen Boyd. 2022. Designing Up with Value-Sensitive Design: Building a Field Guide for Ethical ML Development. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 2069–2082. doi:10.1145/3531146.3534626

[21] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[22] Virginia Braun and Victoria Clarke. 2022. *Thematic analysis: A practical guide*. Sage Publications, Los Angeles, USA.

[23] Virginia Braun and Victoria Clarke. 2023. *Thematic Analysis*. Springer International Publishing, Cham, 7187–7193. doi:10.1007/978-3-031-17299-1_3470

[24] Robin N. Brewer, Christina Harrington, and Courtney Heldreth. 2023. Envisioning Equitable Speech Technologies for Black Older Adults. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 379–388. doi:10.1145/3593013.3594005

[25] Kevin Bryson. 2023. Designing Interfaces to Elicit Data Issues for Data Workers. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 957–958. doi:10.1145/3600211.3604756

[26] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

[27] Patricia Hill Collins. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* (2nd ed.). Routledge, New York.

[28] Patricia Hill Collins and Sirma Bilge. 2020. *Intersectionality*. Polity Press, Cambridge, UK. https://www.wiley.com/en-gb/Intersectionality%2C+2nd+Edition-p-00059152

[29] Combahee River Collective. 1978. A Black Feminist Statement. In *Capitalist Patriarchy and the Case for Socialist Feminism*, Zillah R. Eisenstein (Ed.). Monthly Review Press, New York, 210–218.

[30] Council of Europe, Committee on Artificial Intelligence (CAI). 2024. Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law (HUDERIA Methodology). https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of-arti/1680b2a09f Accessed: 2025-01-22.

[31] Kimberle Crenshaw. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics. *University of Chicago Legal Forum* 1989 (1989), 139–167. Issue 1. http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

[32] Kimberle Crenshaw. 1991. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review* 43, 6 (1991), 1241–1299. http://www.jstor.org/stable/1229039

[33] Angela Y. Davis, Gina Dent, Erica R. Meiners, and Beth E. Richie. 2022. *Abolition. Feminism. Now.* Haymarket Books, Chicago, IL.

[34] Jenny L Davis. 2023. 'Affordances' for Machine Learning. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 324–332. doi:10.1145/3593013.3594000

[35] Jenny L. Davis, Apryl Williams, and Michael W. Yang. 2021. Algorithmic reparation. *Big Data & Society* 8, 2 (July 2021), 20539517211044808. doi:10.1177/20539517211044808 Publisher: SAGE Publications Ltd.

[36] Nathan Dennler, Anaelia Ovalle, Ashwin Singh, Luca Soldaini, Arjun Subramonian, Huy Tu, William Agnew, Avijit Ghosh, Kyra Yee, Irene Font Peradejordi, Zeerak Talat, Mayra Russo, and Jess De Jesus De Pinho Pinhal. 2023. Bound by the Bounty: Collaboratively Shaping Evaluation Processes for Queer AI Harms. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 375–386. doi:10.1145/3600211.3604682

[37] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. MIT Press, Cambridge, MA. https://mitpress.mit.edu/books/data-feminism

[38] Ajay Divakaran, Aparna Sridhar, and Ramya Srinivasan. 2023. Broadening AI Ethics Narratives: An Indic Art View. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 2–11. doi:10.1145/3593013.3593971

[39] Elizabeth Edenberg and Alexandra Wood. 2023. Disambiguating Algorithmic Bias: From Neutrality to Justice. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 691–704. doi:10.1145/3600211.3604695

[40] Amy C. Edmondson. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* 44, 2 (1999), 350–383. doi:10.2307/2666999

[41] Amy C. Edmondson. 2018. *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*. John Wiley & Sons, Hoboken, NJ. https://www.wiley.com/en-us/The+Fearless+Organization:+Creating+Psychological+Safety+in+the+Workplace+for+Learning,+Innovation,+and+Growth-p-9781119477242

[42] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, New York.

[43] Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. 2021. Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 489–503. doi:10.1145/3442188.3445912

[44] James R. Foulds, Rashidul Islam, Kamrun Naher Keya, and Shimei Pan. 2020. An Intersectional Definition of Fairness. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, Dallas, TX, USA, 1918–1921. doi:10.1109/ICDE48307.2020.00203

[45] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. "I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 205–216. doi:10.1145/3593013.3593989

[46] Janneke Gerards, Mirko Tobias Schäfer, Iris Muis, Arthur Vankan, et al. 2022. Fundamental rights and algorithms impact assessment (FRAIA). https://research-portal.uu.nl/en/publications/fundamental-rights-and-algorithms-impact-assessment-fraia Accessed: 2025-05-05.

[47] Marissa Gerchick, Tobi Jegede, Tarak Shah, Ana Gutierrez, Sophie Beiers, Noam Shemtov, Kath Xu, Anjana Samant, and Aaron Horowitz. 2023. The Devil is in the Details: Interrogating Values Embedded in the Allegheny Family Screening Tool. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1292–1310. doi:10.1145/3593013.3594081

[48] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a critical race methodology in algorithmic fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 501–512. doi:10.1145/3351095.3372826

[49] Jacqueline Hannan, Huei-Yen Winnie Chen, and Kenneth Joseph. 2021. Who Gets What, According to Whom? An Analysis of Fairness Perceptions in Service Allocation. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 555–565. doi:10.1145/3461702.3462568

[50] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

[51] Melissa Heikkilä. 2022. Algorithmic Bias in Welfare System: How Automation is Reinforcing Discrimination in Europe. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ Accessed: 2025-05-04.

[52] Amnesty International. 2020. Xenophobic Machines: Discrimination through Unregulated Use of Algorithms in the Dutch Welfare System. https://www.amnesty.org/en/documents/eur35/4686/2021/en/ Accessed: 2025-01-10.

[53] Edward B. Kang. 2023. On the Praxes and Politics of AI Speech Emotion Recognition. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 455–466. doi:10.1145/3593013.3594011

[54] Maximilian Kasy and Rediet Abebe. 2021. Fairness, Equality, and Power in Algorithmic Decision-Making. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 576–586. doi:10.1145/3442188.3445919

[55] Michael Katell, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and P. M. Krafft. 2020. Toward situated interventions for algorithmic equity: lessons from the field. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 45–55. doi:10.1145/3351095.3372874

[56] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 2564–2572. https://proceedings.mlr.press/v80/kearns18a.html

[57] Goda Klumbytė, Claude Draude, and Alex S. Taylor. 2022. Critical Tools for Machine Learning: Working with Intersectional Critical Concepts in Machine Learning Systems Design. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1528–1541. doi:10.1145/3531146.3533207

[58] Bran Knowles, Jasmine Fledderjohann, John T. Richards, and Kush R. Varshney. 2023. Trustworthy AI and the Logics of Intersectional Resistance. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 172–182. doi:10.1145/3593013.3593986

[59] Youjin Kong. 2022. Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 485–494. doi:10.1145/3531146.3533114

[60] Bogdan Kulynych, Rebekah Overdorf, Carmela Troncoso, and Seda Gürses. 2020. POTs: protective optimization technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 177–188. doi:10.1145/3351095.3372853

[61] Susan Leavy, Eugenia Siapera, and Barry O'Sullivan. 2021. Ethical Data Curation for AI: An Approach based on Feminist Epistemology and Critical Theories of Race. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 695–703. doi:10.1145/3461702.3462598

[62] Kevyn Levie. 2021. The Dutch Government's Benefits Scandal Is Rooted in Stigma Against Welfare Recipients. https://jacobin.com/2021/01/dutch-welfare-benefits-childcare-scandal?utm_source=chatgpt.com Accessed: 2025-05-04.

[63] Christina Lu, Jackie Kay, and Kevin McKee. 2022. Subverting machines, fluctuating identities: Re-learning human categorization. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1005–1015. doi:10.1145/3531146.3533161

[64] Alessandro Mantelero. 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review* 54 (2024), 106020. doi:10.1016/j.clsr.2024.106020

[65] Nina Markl. 2022. Language variation and algorithmic bias: understanding algorithmic bias in British English automatic speech recognition. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 521–534. doi:10.1145/3531146.3533117

[66] Nora McDonald and Shimei Pan. 2020. Intersectional AI: A Study of How Information Science Students Think about Ethics and Their Impact. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–19. doi:10.1145/3415218

[67] Morgan Meaker. 2024. Algorithms Policed Welfare Systems For Years. Now They're Under Fire for Bias. https://www.wired.com/2024/10/algorithms-policed-welfare-systems-for-years-now-theyre-under-fire-for-bias/ Accessed: 2025-05-04.

[68] Milagros Miceli, Julian Posada, and Tianling Yang. 2022. Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP, Article 34 (jan 2022), 14 pages. doi:10.1145/3492853

[69] Margaret Mitchell, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. 2020. Diversity and Inclusion Metrics in Subset Selection. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York NY USA, 117–123. doi:10.1145/3375627.3375832

[70] Jared Moore. 2020. Towards a more representative politics in the ethics of computer science. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 414–424. doi:10.1145/3351095.3372854

[71] Giulio Morina, Viktoriia Oliinyk, Julian Waton, Ines Marusic, and Konstantinos Georgatzis. 2020. Auditing and Achieving Intersectional Fairness in Classification Problems. arXiv:1911.01468 [cs.LG] https://arxiv.org/abs/1911.01468

[72] Shiva Omrani Sabbaghi, Robert Wolfe, and Aylin Caliskan. 2023. Evaluating Biased Attitude Associations of Language Models in an Intersectional Context. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 542–553. doi:10.1145/3600211.3604666

[73] Anaelia Ovalle, Arjun Subramonian, Vagrant Gautam, Gilbert Gee, and Kai-Wei Chang. 2023. Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 496–511. doi:10.1145/3600211.3604705

[74] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an Online Infrastructure for Collecting AI Data From People With Disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 52–63. doi:10.1145/3442188.3445870

[75] Gourab K. Patro, Lorenzo Porcaro, Laura Mitchell, Qiuyue Zhang, Meike Zehlike, and Nikhil Garg. 2022. Fair ranking: a critical review, challenges, and future directions. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1929–1942. doi:10.1145/3531146.3533238

[76] Dana Pessach and Erez Shmueli. 2022. A Review on Fairness in Machine Learning. *ACM Comput. Surv.* 55, 3, Article 51 (Feb. 2022), 44 pages. doi:10.1145/3494672

[77] Giada Pistilli, Carlos Muñoz Ferrandis, Yacine Jernite, and Margaret Mitchell. 2023. Stronger Together: on the Articulation of Ethical Charters, Legal Tools, and Technical Documentation in ML. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 343–354. doi:10.1145/3593013.3594002

[78] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J. Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, Hetvi J, Huan Zhang, Jaidev Shriram, Kruno Lehman, Luca Soldaini, Maarten Sap, Marc Peter Deisenroth, Maria Leonor Pacheco, Maria Ryskina, Martin Mundt, Milind Agarwal, Nyx Mclean, Pan Xu, A Pranav, Raj Korpan, Ruchira Ray, Sarah Mathew, Sarthak Arora, St John, Tanvi Anand, Vishakha Agrawal, William Agnew, Yanan Long, Zijie J. Wang, Zeerak Talat, Avijit Ghosh, Nathaniel Dennler, Michael Noseworthy, Sharvani Jha, Emi Baylor, Aditya Joshi, Natalia Y. Bilenko, Andrew Mcnamara, Raphael Gontijo-Lopes, Alex Markham, Evyn Dong, Jackie Kay, Manu Saraswat, Nikhil Vytla, and Luke Stark. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1882–1895. doi:10.1145/3593013.3594134

[79] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York NY USA, 145–151. doi:10.1145/3375627.3375820

[80] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 515–525. doi:10.1145/3442188.3445914

[81] Divya Ramesh, Vaishnav Kameswaran, Ding Wang, and Nithya Sambasivan. 2022. How Platform-User Power Relations Shape Algorithmic Accountability: A Case Study of Instant Loan Platforms and Financially Stressed Users in India. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1917–1928. doi:10.1145/3531146.3533237

[82] Manon Romain, Adrien Senecat, Soizic Pénicaud, Gabriel Geiger, and Justin-Casimir Braun. 2023. How We Investigated France's Mass Profiling Machine. https://www.lighthousereports.com/methodology/how-we-investigated-frances-mass-profiling-machine/ Published December 4, 2023. Co-published with *Le Monde*. Accessed: 2025-05-04.

[83] Manon Romain, Adrien Senecat, Soizic Pénicaud, Gabriel Geiger, Maxime Vaudano, Justin-Casimir Braun, Elsa Delmas, Léa Girardo, Tomas Statius, and Daniel Howden. 2023. France's Digital Inquisition. https://www.lighthousereports.com/investigation/frances-digital-inquisition/ Published December 4, 2023. Co-published with *Le Monde*. Accessed: 2025-05-04.

[84] Loretta J. Ross. 2025. *Calling In: How to Start Making Change with Those You'd Rather Cancel*. Simon & Schuster, New York, NY. https://www.simonandschuster.com/books/Calling-In/Loretta-J-Ross/9781982190798

[85] Loretta J. Ross and Rickie Solinger. 2017. *Reproductive Justice: An Introduction*. University of California Press, Berkeley, CA. https://www.ucpress.edu/book/9780520288201/reproductive-justice

[86] Benita Roth. 2004. *Separate roads to feminism: Black, Chicana, and White feminist movements in America's second wave*. Cambridge University Press, Cambridge, UK ; New York.

[87] Princess Sampson, Ro Encarnacion, and Danaë Metaxa. 2023. Representation, Self-Determination, and Refusal: Queer People's Experiences with Targeted Advertising. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1711–1722. doi:10.1145/3593013.3594110

[88] Ken Schwaber and Mike Beedle. 2001. *Agile Software Development with Scrum* (1st ed.). Prentice Hall PTR, USA.

[89] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems.

In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, Atlanta GA USA, 59–68. doi:10.1145/3287560.3287598

[90] William Seymour, Max Van Kleek, Reuben Binns, and Dave Murray-Rust. 2022. Respect as a Lens for the Design of AI Systems. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society.* ACM, Oxford United Kingdom, 641–652. doi:10.1145/3514094.3534186

[91] Anastasia Siapka. 2022. Towards a Feminist Metaethics of AI. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society.* ACM, Oxford United Kingdom, 665–674. doi:10.1145/3514094.3534197

[92] Mona Sloane and Janina Zakrzewski. 2022. German AI Start-Ups and "AI Ethics": Using A Social Practice Lens for Assessing and Implementing Socio-Technical Innovation. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Seoul Republic of Korea, 935–947. doi:10.1145/3531146.3533156

[93] Jessie J. Smith, Anas Buhayh, Anushka Kathait, Pradeep Ragothaman, Nicholas Mattei, Robin Burke, and Amy Voida. 2023. The Many Faces of Fairness: Exploring the Institutional Logics of Multistakeholder Microlending Recommendation. In *2023 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Chicago IL USA, 1652–1663. doi:10.1145/3593013.3594106

[94] Wonyoung So, Pranay Lohia, Rakesh Pimplikar, A.E. Hosoi, and Catherine D'Ignazio. 2022. Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Seoul Republic of Korea, 988–1004. doi:10.1145/3531146.3533160

[95] KIMBERLY SPRINGER. 2005. *Living for the Revolution: Black Feminist Organizations, 1968–1980.* Duke University Press, Durham, NC. http://www.jstor.org/stable/j.ctv120qt2c

[96] Harini Suresh, Rajiv Movva, Amelia Lee Dogan, Rahul Bhargava, Isadora Cruxen, Angeles Martinez Cuba, Guilia Taurino, Wonyoung So, and Catherine D'Ignazio. 2022. Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Seoul Republic of Korea, 667–678. doi:10.1145/3531146.3533132

[97] Jacob Thebault-Spieker, Sukrit Venkatagiri, Naomi Mine, and Kurt Luther. 2023. Diverse Perspectives Can Mitigate Political Bias in Crowdsourced Content Moderation. In *2023 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Chicago IL USA, 1280–1291. doi:10.1145/3593013.3594080

[98] Anna-Lena Theus. 2023. Striving for Affirmative Algorithmic Futures: How the Social Sciences can Promote more Equitable and Just Algorithmic System Design. In *2023 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Chicago IL USA, 558–568. doi:10.1145/3593013.3594022

[99] Jakita O. Thomas, Neha Kumar, Alexandra To, Quincy Brown, and Yolanda A. Rankin. 2021. Discovering intersectionality: part 2: reclaiming our time. *Interactions* 28, 4 (July 2021), 72–75. doi:10.1145/3468783

[100] Nenad Tomasev, Kevin R. McKee, Jackie Kay, and Shakir Mohamed. 2021. Fairness for Unobserved Characteristics: Insights from Technological Impacts on Queer Communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society.* ACM, Virtual Event USA, 254–265. doi:10.1145/3461702.3462540

[101] European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng Accessed: 2025-01-13.

[102] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. Data Leverage: A Framework for Empowering the Public in its Relationship with Technology Companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Virtual Event Canada, 215–227. doi:10.1145/3442188.3445885

[103] Gloria Wekker. 2016. *White Innocence: Paradoxes of Colonialism and Race.* Duke University Press, Durham, NC.

[104] Eline Westra. 2024. Multiple barriers to the Dutch welfare state. Black Feminists' intersectional claims to social citizenship in the 1980s. *Critical Social Policy* 44, 3 (2024), 403–424. https://doi.org/10.1177/02610183231215232 Accessed: 2024-11-20.

[105] Rüdiger Wirth and Jochen Hipp. 2000. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, Vol. 1. Manchester, UK, 29–39.

[106] Robert Wolfe, Mahzarin R. Banaji, and Aylin Caliskan. 2022. Evidence for Hypodescent in Visual Semantic AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Seoul Republic of Korea, 1293–1304. doi:10.1145/3531146.3533185

[107] Robert Wolfe and Aylin Caliskan. 2022. Markedness in Visual Semantic AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Seoul Republic of Korea, 1269–1279. doi:10.1145/3531146.3533183

[108] Robert Wolfe, Yiwei Yang, Bill Howe, and Aylin Caliskan. 2023. Contrastive Language-Vision AI Models Pretrained on Web-Scraped Multimodal Data Exhibit Sexual Objectification Bias. In *2023 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Chicago IL USA, 1174–1185. doi:10.1145/3593013.3594072

[109] Raphaële Xenidis. 2022. Intersectionality from Critique to Practice: Towards An Intersectional Discrimination Test in the Context of 'Neutral Dress Codes'. *European Equality Law Review* 2 (2022), 17 pages.

[110] Iris Marion Young. 2011. *Responsibility for justice.* Oxford University Press, New York.

[111] Miri Zilka, Riccardo Fogliato, Jiri Hron, Bradley Butcher, Carolyn Ashurst, and Adrian Weller. 2023. The Progression of Disparities within the Criminal Justice System: Differential Enforcement and Risk Assessment Instruments. In *2023 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Chicago IL USA, 1553–1569. doi:10.1145/3593013.3594099

A Actionable Recommendations per Theme

An overview of the actionable recommendations is presented in Table 2, each one linked to one of the five themes.

B Thematic Analysis: Six Phase Method

As written in section 3.1.2, we have chosen the inductive reflexive TA Braun and Clarke [21]. To provide structure to the flexible method, Braun and Clarke [21] defined a process of six phases. We highlight each phase below (with the name in italics), whilst demonstrating our implementation.

For phase one, *familiarizing yourself with the dataset*, the authors of this work read all the abstracts of the selected papers and divided the papers between them to read fully.

Phase two is *coding*. To generate the initial codes, we started with a set of five papers per author where we read the whole paper and selected (copied) paragraphs in a table if they contained at least one recommendation. Then, for each paragraph containing a recommendation, the recommendation was coded by summarising and/or rephrasing it to its essence. This process was fine tuned iteratively, such that the various readers would code in a similar style yet influenced by the different backgrounds. After the first iteration, we decided that we would extract coding of papers by reading the introduction fully and then deciding which sections of the papers needed to be read to find the recommendations. We extracted 206 recommendations/codes.

Phase three, *generating initial themes.* With a finished set of codes, we searched for themes. The codes were grouped with identical or similar codes and collected in thematic groups. In the initial round, 17 themes were found. In following iterations the themes were further grouped together to finally reach the five themes presented in this paper, see Table 3.

Phase four, *developing and reviewing themes.* Through an iterative process on whether adjusting the themes and reassigning codes to different themes, as well as checking whether the themes encompass the codes, we grouped the themes further into finally reaching five overarching themes. Additionally, each of us picked two to four most relevant papers of their selection which was then read by all authors to enhance a common understanding. This also supported verification that the five chosen themes spanned the recommendation of (most of) our selected papers.

Phase five, *refining, defining and naming*. In this phase, we first organized brainstorm sessions between the authors to refine, name and define the themes with our goal of actionable recommendations. In this fifth phase, a first iteration of the five refined, defined and

Table 2: Actionable Recommendations per Theme

Theme	Actionable Recommendation
Collaboration & Role	Collaborate with multiple disciplines before going into technical details. Use central role of AI experts to invite other disciplines and share responsibilities. Dedicate time and effort to create a psychologically safe environment.
Position and Reflection	Write a positionality statement and reflect on it. Document perspectives and decisions throughout the lifecycle of AI.
Community and Participation	Invite communities to co-own the participation process. Make participation financially sustainable for communities. Design a mechanism where impacted communities can safely voice concerns.
Power and Social Context	Position the AI within social context and define the present power relations. Redefine concepts with power and social context in mind.
Data and Metrics	Be critical on your objective with data and metrics. Augment quantitative approaches with qualitative research and participatory design. Document clearly on the intended use and limitations of data, model and metrics.

Table 3: Overview of 17 themes grouped into 5 recommendation areas

Collaboration and Role 1. Need for interdisciplinary teams/knowledge 2. Responsibility AI expert 3. Common language
Position and Reflection 4. Reflective 5. Positionality
Community and Participation 6. Participatory design 7. Value elicitation
Power and Social Context 8. Framing (justice, power) 9. Beyond status quo/social context 10. AI value chain power dynamics 11. Policy
Measurement and Nuance *(now Data and Metrics)* 12. Forms of bias/harms 13. Metrics 14. Inclusive data collection
Other 15. Ethical review 16. Pro-active/early 17. Teaching

named themes were evaluated with an interactive workshop with a diverse set of AI experts, see section 5. Based on their input we refined, and renamed the themes and sub-themes such that they are engaging and informative.

Finally, phase six of *writing up*, constituted the reporting of our themes in this paper. As part of the analytical process, the themes are refined further through the act of writing, positing the work in other scholarship and particularly our structure of actionable recommendations.

C Selected Papers

Table 4 shows all the papers selected for the literature survey.

Table 4: Overview of papers for literature, sorted by author names

Authors	Conference
Albert & Delano [1]	ACM FAccT 2023
Aler Tubella et al. [2]	ACM FAccT 2023
Barlas et al. [6]	ACM AIES 2021
Barnett & Diakopoulos [7]	ACM AIES 2022
Bates et al. [8]	ACM FAT* 2020
Benbouzid [9]	ACM FAccT 2023
Bender et al. [10]	ACM FAccT 2021
Bennett et al. [11]	ACM AIES 2023
Benthall & Haynes [12]	ACM FAT* 2019
Bergman et al. [14]	ACM FAccT 2023
Birhane et al. [15]	ACM FAccT 2022
Black et al. [16]	ACM FAccT 2022
Blili-Hamelin & Hancox-Li [17]	ACM FAccT 2023
Boag et al. [18]	ACM FaccT 2022
Bondi et al. [19]	ACM AIES 2021
Boyd [20]	ACM FAccT 2022
Brewer et al. [24]	ACM FAccT 2023
Bryson [25]	ACM AIES 2023
Davis [34]	ACM FAccT 2023
Dennler et al. [36]	ACM AIES 2023
Divakaran et al. [38]	ACM FAccT 2023
Edenberg & Wood [39]	ACM AIES 2023
Finocchiaro et al. [43]	ACM FAccT 2021
Gadiraju et al. [45]	ACM FAccT 2023
Gerchick et al. [47]	ACM FAccT 2023
Hanna et al. [48]	ACM FAT* 2020
Hannan et al. [49]	ACM AIES 2023
Kang [53]	ACM FAccT 2023
Kasy & Abebe [54]	ACM FAccT 2021
Katell et al. [55]	ACM FAT* 2020
Klumbytė et al. [57]	ACM FAccT 2022
Knowles et al. [58]	ACM FAccT 2023
Kulynych et al. [60]	ACM FAT* 2020
Leavy et al. [61]	ACM AIES 2021
Lu et al. [63]	ACM FAccT 2022
Markl [65]	ACM FAccT 2022
Mitchell et al. [69]	ACM AIES 2020
Moore [70]	ACM FAT* 2020
Omrani Sabbaghi et al. [72]	ACM AIES 2023
Ovalle et al. [73]	ACM AIES 2023
Park et al. [74]	ACM FAccT 2021
Patro et al. [75]	ACM FAccT 2022
Pistilli et al. [77]	ACM FAccT 2023
Queerinai et al. [78]	ACM FAccT 2023
Raji et al. [79]	ACM AIES 2020
Raji et al. [80]	ACM FAccT 2020
Ramesh et al. [81]	ACM FAccT 2022
Sampson et al. [87]	ACM FAccT 2023
Selbst et al. [89]	ACM FAT* 2019
Seymour et al. [90]	ACM AIES 2022
Siapka [91]	ACM AIES 2022
Sloane & Zakrzewski [92]	ACM FAccT 2023
Smith et al. [93]	ACM FAccT 2023
So et al. [94]	ACM FAccT 2022
Suresh et al. [96]	ACM FAccT 2022
Thebault-Spieker et al. [97]	ACM FAccT 2023
Theus [98]	ACM FAccT 2023
Tomasev et al. [100]	ACM AIES 2021
Vincent et al. [102]	ACM FAccT 2021
Wolfe & Caliskan [107]	ACM FAccT 2022
Wolfe et al. [106]	ACM FAccT 2022
Wolfe et al. [108]	ACM FAccT 2023
Zilka et al. [111]	ACM FAccT 2023