Participatory AI Justice in HCI: A Scoping Review

Maria Luce Lupetti
Department of Architecture and Design
Politecnico di Torino
Turin, Italy
maria.lupetti@polito.it

Cristina Zaga
Human-Centred Design
University of Twente
Enschede, NY, Netherlands
c.zaga@utwente.nl

Nazli Cila
Human-Centered Design
Delft University of Technology
Delft, Netherlands
n.cila@tudelft.nl

Abstract

Participatory design is increasingly used to address the negative social impacts of artificial intelligence (AI), aiming for more inclusive and equitable innovation. However, it can inadvertently reproduce injustice and reinforce power imbalances, even with good intentions. While the HCI community is critical of these issues, the existing knowledge is often fragmented, making it challenging for AI researchers and policymakers to navigate. This paper presents a scoping review of participatory AI research in HCI focused on justice. We detail how participatory AI unfolds in practice and offers methodological insights on the roles of researchers and partnership with communities, the practical and contextual challenges, the role of reflexivity and situatedness and the essential but not so central role of artefacts in participatory processes. We conclude with recommendations for engaging in participatory design to promote justice in AI systems.

CCS Concepts

• **Do Not Use This Code Generate the Correct Terms for Your Paper**;

Keywords

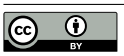
Participatory AI, Participatory Design, Social Justice, Design Justice, Artificial Intelligence

ACM Reference Format:
Maria Luce Lupetti, Cristina Zaga, and Nazli Cila. 2026. Participatory AI Justice in HCI: A Scoping Review. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3772318.3791704

1 Introduction

> *"It is not coincidental that participatory design offers these challenges at a time when many businesses are discarding the conventional wisdom and struggling to remake themselves in the face of heightened international competition, and dramatic political change is taking place around the globe"* [79]

It was 1993 when Muller and Kuhn [79] mentioned this quote in their seminal article published by Communications of the ACM. They were witnessing times of significant change, both in corporate environments and the political landscape, that were opening up avenues for design to play a different role: that of a mediator, or even an advocate, for the needs of people for whom artefacts were being designed. One could argue that, despite decades of participatory design work and a steady evolution of methods and theories around it, not much has changed. Today, more than ever, participatory design remains a key human-computer interaction (HCI) approach to help navigate the muddy land that corporations and institutions shape, and that communities must walk through. We live once again in times of severe political instability, with wars taking place in areas from both the global south and the west [100], right-wing and conservative parties gaining momentum in more and more countries [14], our environment showing signs of collapse [81], and poverty rising globally [80]. Countering all this, there is an inebriating spring of Artificial Intelligence (AI) development blooming globally. The number of large-scale AI models released has significantly increased since the debut of GPT-2 in 2019. By 2024, 167 large-scale models were launched [50], integrating AI into various applications. These advancements are transforming access to resources, labour dynamics, and civic discourse, but they also present significant global challenges. As McQuillian ([77] page 10) underscores, AI is an apparatus deeply entangled with our social, political, economic, and environmental infrastructures. To give a few examples, the widespread adoption of AI is seen to be causing a surge in emissions and energy consumption, especially for tech giants [82], while also causing drought in areas that are already suffering from water scarcity [85]. AI-generated deep-fakes increasingly enter political elections and steer the public sentiment through misinformation [109]. In a similar vein, far-right parties often use AI-generated content to support populist narratives, e.g., anti-immigrant rhetoric [90]. AI-based decision support systems are showing a dramatic impact of algorithms on warfare, causing a dramatic increase in violence against civilians and an erosion of human accountability (see the current use of AI by the Israeli military force in Gaza [35]). And last but not least, the adoption of AI in various public services and systems is manifesting issues of systematic inequality and biased treatment and distribution of resources. For example, the adoption of machine learning technologies in mortgage application processes carries risks for black and white Hispanic borrowers, who are disproportionately less likely to gain credit [45]. Similarly, research has shown that the use of algorithmic diagnosis in chest radiographs can consistently and selectively underdiagnose underserved patient populations, especially those of intersectional underserved sub-populations, such as Hispanic female patients [93].



CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791704

This can be shortened like this: AI developments are closely linked to social, economic, environmental, and political issues. As AI is implemented, it is crucial to consider who is affected, how, when, and where this impact occurs [104]. Concerns about uncertainties, lack of human oversight, and unintended consequences are particularly important when AI is used for critical tasks like evaluating financial trustworthiness or making medical diagnoses. Given the current context, *participatory design practices are gaining momentum again.* The idea is that involving the public, especially systematically excluded communities, in the design process will ultimately lead to the creation of AI systems that positively impact society.[5, 17, 19, 91, 115]. In fact, along with technical strategies to ensure more robust and accountable AI solutions, the critical AI community is calling for participatory AI (PAI) to be integrated into AI development practices [17]) – a participatory turn designed to ensure that human values and rights become actionable design requirements for the development of AI systems [5, 10, 114] and that moral boundaries are respected [24]. This call also aligns with the recent views and strategies of numerous international institutions, such as the European Union's special interest group on AI Ethics, as highlighted in the *Ethics Guidelines for Trustworthy AI* report. [57] argues for the participation of relevant stakeholders as a key practice to ensure diversity, nondiscrimination, fairness, and general acceptability of AI models and implementations. Similarly, the Organisation for Economic Cooperation and Development (OECD) argues that the implementation of AI in public sector institutions should come through an integrated approach that incorporates participatory mechanisms as a way to ensure alignment with public values and objectives [110].

However, as design scholarship warns us, participation practices are inherently subject to inclusivity issues and power struggles [10, 95]. Despite its long tradition and growth in methods and scopes, participatory design consistently fails to understand and address how it perpetuates historically imbalanced power dynamics [10, 15, 18], and how it leaves unattended the radical issues that marginalized communities experience daily (i.e., complexities associated with the location they live in, language and knowledge barriers, financial burdens, time and more [38]) and that hinder their inclusion in design processes [10, 54, 95]. The growing popularity of these practices in AI innovation contexts is likely to exacerbate these issues. The participatory ambition to be inclusive and open to public deliberation clashes with the configuration and dynamics of the current AI landscape, which is characterised by *centrally-owned-yet-globalised* infrastructures governed by tech giants with explicit capitalist ambitions [117]. The focus on localised knowledge from community engagement contrasts with the goal of foundation models to be versatile across different domains and regions. [103]. Moreover, the resources and time needed to engage with communities clash with the pace of technological innovation, which cannot afford the practical implications of community participation.[19, 29]. As a result, the majority of PAI efforts end up being performative. Stakeholders are largely involved to provide input on discrete implementation parameters, rather than making key AI design decisions [29]. Experts do not even take the needs or recommendations of the different stakeholder groups into full consideration [95]. Participatory engagements are primarily utilised to improve AI systems and are often exploited by corporate entities to maximise profits.[17].

Critical HCI and AI scholarship is increasingly aware of these issues. Research is directly addressing matters of social justice in the participatory design of AI systems, with studies trying to involve communities throughout the whole AI development lifecycle [87, 116] and critical reviews helping to build literacy on how matters of justice are entangled with participation processes [26, 29, 41]. Yet, enormous variations remain in terms of methods and theories applied to achieve participation (Delgado et al., 2023), and access to these resources remains scattered. It is still challenging, especially for AI practitioners –developers and regulating bodies alike– to find and use the body of knowledge that the HCI community, and the critical AI subcommunity in particular, produces. We take a similar stance: having to initiate applied research on PAI and social justice, we find ourselves lacking blueprints for how to approach this space, and we are searching for a holistic resource that can help us address the various dimensions that must be considered when conducting participatory AI. Therefore, we aim to provide the HCI/critical AI community with *a comprehensive view of how matters of justice and participatory AI intersect.*

We present a scoping literature review that examines articles published in leading HCI conferences and journals. This review highlights and critically reflects on contributions that adopt critical perspectives on AI and emphasise issues related to social justice. The goal is not to highlight gaps in the literature, but rather to acknowledge the value of what is already being produced, to learn from good practices. In the end, we establish a foundation for future PAI practices rooted in previous examples that demonstrate how participation in AI can occur with a clear commitment to social justice. The aim of this work, then, is to celebrate, connect and holistically present key knowledge that our community is producing in this space. As such, we contribute with 1) a holistic understanding of participatory AI practices through the lens of social justice; 2) a set of methodological reflections for practising justice-oriented PAI work; and 3) a discussion of future research opportunities.

2 Related Works

The AI community is increasingly embracing participatory practices to develop and implement AI systems that positively impact society [74]. These practices aim to incorporate the needs, perspectives, and concerns of the broader public, including communities that are systematically excluded, into the design process [17, 19]. PAI is an approach to increase the social responsibility of AI and reduce bias both in the logic and data through a collaborative decision-making process [40]. It is regarded as a powerful mitigation strategy for the increasing emergence of adverse impacts that AI can have on society [119]. As Birhane and colleagues [17] argue, "*participation is meant to move beyond individual opinion to centre the values of inclusion, plurality, collective safety, and ownership, subsequently shifting the relationship from one of designer-and-user to one of co-designers and co-creators*".

PAI plays out in different formats and moments of the AI development lifecycle. PAI practices range from public debates with non-expert voices to anticipate risks and potential harms, to techniques to improve the design of technical systems by involving

diverse stakeholders in the design process, as well as approaches in between [51]. The AI community (engineers, designers, industry professionals, HCI researchers and more) unanimously considers PAI as an approach that should continue to grow and be refined as a key component of the AI development and deployment lifecycle as a way for communities, especially those at the margins of society, to contribute to strategic decisions around AI [17]. However, PAI is a field with inherent complexities and contradictions [74]; some are inherited from the very participatory design tradition it stems from, and some are peculiar to AI systems development. In what follows, we unpack the risks and issues of participation, both in the participatory design tradition and in the current critical AI literature.

2.1 Participatory Design and its social justice issues

Participatory Design (PD) is an approach that "*questions major assumptions about technologies in workplaces, communities, homes, and social institutions*" [79]. Blending methodological traditions of related fields, such as science and technology studies, sociology, and psychology, with design expertise, PD makes "*exploratory and speculative tools, services, and products in the form of worksheets, prototypes, representations, event schedules, gatherings, mock-ups, and other design things*" to collaboratively envision preferable futures [71]. It opens the innovation process to a joint effort between people from different organisations and backgrounds, up to the point of sharing and combining ideas – what we usually refer to as *co-design* [101].

The origin of PD lay in the Scandinavian workplace democracy movements, and historically grew within the space of critical approaches to work technology development, especially computer-based systems, which made evident, over time, matters of power distribution in the workplace [65]. PD was initially motivated by practical matters–a need for gathering experiential knowledge about artefacts in use, for which it examines "*the tacit, invisible aspects of human activity*" [98], with the ultimate intent of developing products that better align with user needs and skills [59]. Participatory design (PD) has gained popularity due to its practical value and is now used in various fields, including entertainment technologies, urban planning, assistive robotics, and automated vehicles. It highlights not only utilitarian benefits but also the socio-political implications of the design process[79], and an ambition to dismantle the power imbalances between "*those charged with technology design and those who must live with its consequences*" [65]. As such, PD has increasingly been regarded as a practice that is "*antithetical to consumerism*" [92] and that can encourage a view of design as productive dialogue and cooperation among diverse stakeholders, potentially holding contrasting values. This led to a gradual shift in focus from individual projects to broader participation and development, evolving into a tool for organisational change and influencing legal and political frameworks. [65].

In this vein, more and more work is now "*leveraging participatory design methods to engage with marginalised and underserved populations given the association of empowerment and, in some cases, activism that may result from this engagement*" [54]. Participatory design (PD) is increasingly adopted in community-centred movements like the Design Justice Network due to its potential to influence social change. The network aims to harness "*collective power and experiences to bring forth safer, more just, more accessible, and more sustainable*" worlds [34]. As the operational space of PD expands, as reflected in its enriched nomenclatures like Community-based Participatory Design [32], so do its methodological concerns and ambitions. From initial considerations of when to employ these practices, who to engage, and how many at the same time [79], now PD has become a space where researchers and practitioners question their own actions and identities, to achieve equity and justice. Designers reflect on their privileges and positions of power [54, 118], and **the very act of designing is increasingly recognised as a universal human activity** [27]. We are witnessing a radical rethinking–even a downsizing– of the role of the designer, who is better seen as a mediator and facilitator rather than a decision-maker [33, 54, 92], and a shift in innovation practices, one that questions the centrality of technological progress up to the point of being open to the possibility of technology refusal [88] and resistance [4].

Despite these noble intentions, PD often fails to enact meaningful participation and continues to grapple with the very justice issues it aims to address. As Lodato and DiSalvo [71] argue, especially when applied in the public realm, PD is performed within invisible but strict constraints. These include differing ideals and values of stakeholders, inadequate administrative and legal frameworks to support the envisioned possibilities, and practical limitations on the range of possible experiments. At the time of participatory engagements, *design decisions are often already taken*: the *actual* design happens elsewhere at other times, i.e., when a technology is developed. The PD processes end up being actions for smoothing AI systems deployment [71]. Understanding these constraints is crucial as they shape design processes, highlighting how design can perpetuate inequalities and power structures. [25]. Inequalities, in fact, manifest at all levels of the design process [27] and are particularly exacerbated by a specific design culture: one that stems from a privileged, white, youthful, and upper-to-middle-class approach to innovation [54]. In this regard, Harrington and colleagues [54] illustrate how "*design workshops are a socially and culturally constructed practice that brings with it expectations that may further marginalise and ultimately undermine the participation of certain individuals*", especially because these are not backed up by a thorough understanding the historical context these communities live in and are generated from. As a consequence, PD often devalues and reduces the lived experiences of those who may not fit the 'norms' of participants in a dominant design culture, and can draw further distinctions between participants and researchers [54]. This adds up to the already acknowledged complexity of involving people in ways that are most conducive to their ability to participate [92], and the need for finding ways to give voice to everyone's lived experience as unique and brilliant contributions to design [33].

Therefore, it remains an open challenge for designers to practice PD equitably [96], especially when engaging underserved communities, and to decolonise these practices [54]. Designers must interrogate themselves about their positions of power and about how to practice design "*to sustain, heal, and empower communities, as well as to seek liberation from exploitative and oppressive systems*"

[33], which, according to Harrington and colleagues [54], must start with accounting for the history and context of where participation happens as precondition for building trustworthy processes that challenge corporate design thinking traditions and are truly open to the richness and fullness of participants stories about their lived experiences, views and aspirations.

2.2 Current critiques to Participatory AI

As seen in other sectors where PD has been applied, PAI initiatives carry the risk of superficial engagements in which communities are involved only in **tokenistic ways.** [89]. These initiatives stem from positive intentions but often result in "participation washing" instead of fostering meaningful engagement. [39, 89, 95]. The underlying causes are complex. There is a tension between the desire for inclusivity in public deliberation and the current AI landscape, which is dominated by centralised infrastructures owned by tech giants with capitalist goals. [117]. Secondly, there is a tension between the PD's commitment to localised and context-specific knowledge gained from engaging with communities and the aspiration for foundation models to be versatile and applicable across various domains and geographies.[103]. These tensions add to two existing issues: the practical challenges of practising participatory AI, which requires resources and time that races for technological innovation generally cannot afford [19, 29], and the extractive nature of AI itself.

As Sloane and colleagues argue [94], the field is "*deeply participatory since AI is—by design—only possible through data production by ordinary technology users and their continuous optimisation of AI systems*", but this form of participation is neither representative nor consultative; it is a form of hidden and poorly paid work. AI development involves not just scientists and engineers but also vital roles like annotators, auditors, and testers. These individuals help in data preparation, product review, and model performance assessment. [3]. Furthermore, as models continue to scale by getting trained on ever larger swaths of human-generated data and "fine-tuned" through processes involving human feedback in some form, the risk increases that participation will be reduced to a mere optimisation exercise. This stands in contrast to the ideal of participation as a way for fostering meaningful engagement, empowering broader groups of stakeholders to have a say in AI development [29]. Critical AI studies scholars have sought to address this issue, for example, by acknowledging and emphasising the role of annotators as bearers of situated knowledge, which can be critical to achieve contextually relevant value alignment of large language models [8, 66]. However, when participation takes the form of work [95], possibilities for meaningful contribution remain limited, as such work is governed by overarching infrastructures and politics over which workers have no control. As a direct consequence, AI comes with an inherent barrier for the public to genuinely contribute to its shaping, which Suresh and colleagues [103] refer to as the participatory ceiling.

It needs to be noted, however, that while meaningful participation is inextricably challenging when it comes through data work [78], it may be less so in other forms of participation. Accordingly, scholars argue for the need to understand that PAI comes in various forms at different stages of the AI development lifecycle. As Corbett and colleagues [26] argue, most contemporary AI systems are developed through a process that typically comprises four interconnected steps, i.e., problem formulation, dataset development, model design and training, deployment and monitoring, and each of these may afford some entry points for people to participate, as well as larger or smaller margins for people to influence decision making around the developed AI system.

Sloane and colleagues [95] distinguish three broad modes of participation in AI. The first we mentioned above, participation as work, is surely the wider form of participation, yet it contrasts with the PD tradition and ambitions of democratic deliberation and contestation. Closer to PD tradition, instead, are the two other modes: participation as consultation and participation as justice. In participation as consultation, designers and technologists involve stakeholders through short-term engagements, often workshops or online dialogic sessions, and focus on specific aspects of the system under development. Participation as justice, instead, usually unfolds as long-term, collaborative partnerships with diverse stakeholders to address structural harms and promote equity, often by explicitly building on social justice, crip technoscience, intersectional feminism, and decolonial theories. These practices prioritise "*designing with*" marginalised communities throughout the whole development lifecycle, as a way to ensure that outcomes hold social and political value. In a similar vein, other scholars discussed the various ways participation can be practised. Examples include the Ada Lovelace Institute's five-level participatory mechanism, ranging from inform, consult, involve, and collaborate to empower (by the Ada Lovelace Institute, as cited by [29]), and Corbett et al.'s [26] adaptation of Sherry Arnstein's Ladder of Citizen Participation [7]. The latter presents an eight-rung ladder, where each rung depicts the degree of power afforded to people, spanning from nonparticipation, through tokenism, to citizens' power.

These forms of participation are all possible and do not exclude each other. However, genuine modes of engagement and the space of 'participation as justice, citizens' control, and empowerment' are particularly challenging as AI is developed in and for corporate or public-sector contexts, which are inherently extractive and oppressive. PAI builds on centuries of systemic and institutionalised injustice and inequity, especially in the technology development sector, and (in most cases) fails to acknowledge how AI development adds burdens on communities that consistently experience discrimination, stigmatisation, and inequity [89]. It remains difficult for designers and engineers conducting participatory AI to account for how historical mechanisms of oppression have shaped design spaces and practices [26]. As Birhane and colleagues [17] remind us, "*When we invite people to participate, it is never everyone*": the ones who are typically excluded, those with low literacy, little time, and living at the margins, never gain a seat at the PAI table, let alone a voice. This systematic exclusion is further exacerbated by the widespread belief, even among experts, that AI algorithms are too complex to be meaningfully explained, leaving non-experts without the instruments needed for proper deliberation [39, 94].

Feminist and decolonial AI critiques, however, warn us about the fallacies of this thinking and argue that the very problem lies with "*AI's epistemological apartheid*" ([77] page 110), which, under the mirage of objectivity and scientism, neglects alternative and essential "*ways of knowing, rooted in the lived experience of people*

who are marginalized or minoritized" and that could help us see "*the ways that the assumptions, discursive frameworks and conceptual schemes generated by dominant groups get hard coded into the ways the rest of us think about the natural world and about social relations*" ([77] page 105). It is imperative, then, for the critical AI community to find ways to challenge AI scientism, to dismantle public perceptions of AI as a magical entity [73, 94], and to empower marginalised voices to tell alternative stories of AI [94]. More than AI literacy and expertise, what is needed is a co-defined language and pace for participatory AI to thrive and genuinely empower communities. Researchers, designers and engineers should invest time and resources in building a relationship with the involved communities, where terms for consent [26] and modes of participation [17] are negotiated. Meaningful participation demands long-term presence in AI research, institutions and practices [89] as well as dedicated infrastructures, forms of organised bodies that situate power outside of the public institutions [26], maybe in the fashion of what McQuillian envisions as *people's councils* – a grouping of those affected by AI that form a directly democratic body to decide what to do about it ([77] page 127) – or more concretely, as what we are witnessing with bottom up tech workers initiatives, like *Tech Workers Coalition* and *African Tech Workers Rising*, just to name a couple.

Delegating AI contestation and societal alignment to community-initiated coalitions, however, is not sufficient. Researchers, designers and engineers in the field of HCI, and the critical AI subcommunity more precisely, are increasingly aware of all these complexities and needs, and the responsibilities they bear. Yet, most of the PAI efforts end up being performative. Stakeholders tend to be involved for providing input on discrete implementation parameters, rather than to be empowered to make key AI design decisions [29], and experts very often do not even take the needs or recommendations of the different stakeholder groups into consideration [95]. Participants' engagement often serves as an aid in the refinement of AI systems; efforts are capitalised by corporate actors to maximise profit [17], and participation becomes once again a form of corporate co-optation [17].

Many of these issues stem from the capitalist infrastructure that AI "lives in" and the systemic problems of power and oppression that our society suffers from. Yet, we as researchers, designers and engineers practising participatory AI have a role to play; we can and must strive for justice when performing participatory AI. The problem, however, is that even when knowledge is present, this may be hard to access, especially in its comprehensive form. We find precious literature with scattered pieces of helpful knowledge and recommendations, but we lack a holistic view of how to practice participatory AI in a socially just manner. Over the last years, scholars have given us "*better clarity on what participation is, who it is supposed to serve, how it can be used in the specific context of AI*" [17] and provided us with ways to compare and contrast how different approaches can affect the power afforded to people, such as the use of the Ladder of Citizen Participation [26].

However, significant variations still exist in the methods and theories used to achieve participation, and access to these remains inconsistent [29]. Hence, it is essential to evaluate and understand the current PAI landscape.

3 Method

The goal of this work is to understand how current PAI work engages with justice and whether it offers practical resources to put justice principles in action. With this objective, we conducted a scoping literature review [6] of articles published in leading HCI conferences and journals. Following Arksey and O'Malley's [6] framework, our review process involved: (i) identifying the goal of the review, (ii) identifying relevant studies by defining data sources, (iii) describing the corpus of studies selected, and (iv) conducting a deductive thematic analysis using an iteratively developed codebook. Furthermore, in line with feminist HCI scholarship [13], we acknowledge the interpretative and situated nature of our approach [55] and therefore articulate the positionalities that informed our investigation and the standpoint from which we approach the current landscape of critical AI literature.

3.1 Positionality

As three Assistant Professors working across design, HCI, and AI, we position ourselves within feminist, critical, and justice-oriented traditions in technology research. Commitments to diversity, equity, inclusion, and the interrogation of power in sociotechnical systems shape our shared standpoint. One of us approaches this work through militant design and critical AI, motivated by both personal passion and the urgent need to resource participatory AI projects. Another grounds their scholarship in posthuman feminism and activist practice, developing transdisciplinary methods to reimagine AI and robotics alongside workers, unions, and civic actors. The third brings a design research perspective attentive to democracy, accountability, and the methodological infrastructures that govern participation in AI. Together, we seek to build reference points and frameworks that not only advance academic debates but also sustain collective practices of resistance, reappropriation, and reimagining in human-technology relations.

3.2 Identifying Data Sources and Corpus Selection

All three authors collaborated closely and iteratively to determine the data sources and finalise the selection of the corpus of articles.

3.2.1 Databases, Conferences and Journals. The review was conducted in the Association for Computing Machinery Digital Library (**ACM DL Library**), as it indexes the primary venues in HCI and related fields where critical perspectives on AI are published.

We targeted conferences and journals with a demonstrated dedication to the matters of social justice, including the ACM Conference on Human Factors in Computing Systems (**CHI**), the ACM Designing Interactive Systems Conference (**DIS**), the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (**AIES**), the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (**EAAMO**), the ACM Conference on Fairness, Accountability, and Transparency (**FAACT**), the ACM Conference on Computer-Supported Cooperative Work & Social Computing (**CSCW**), the ACM/IEEE International Conference on Human-Robot Interaction (HRI), the ACM Participatory Design Conference (**PDC**), the ACM Transactions on Computer-Human Interaction Journal

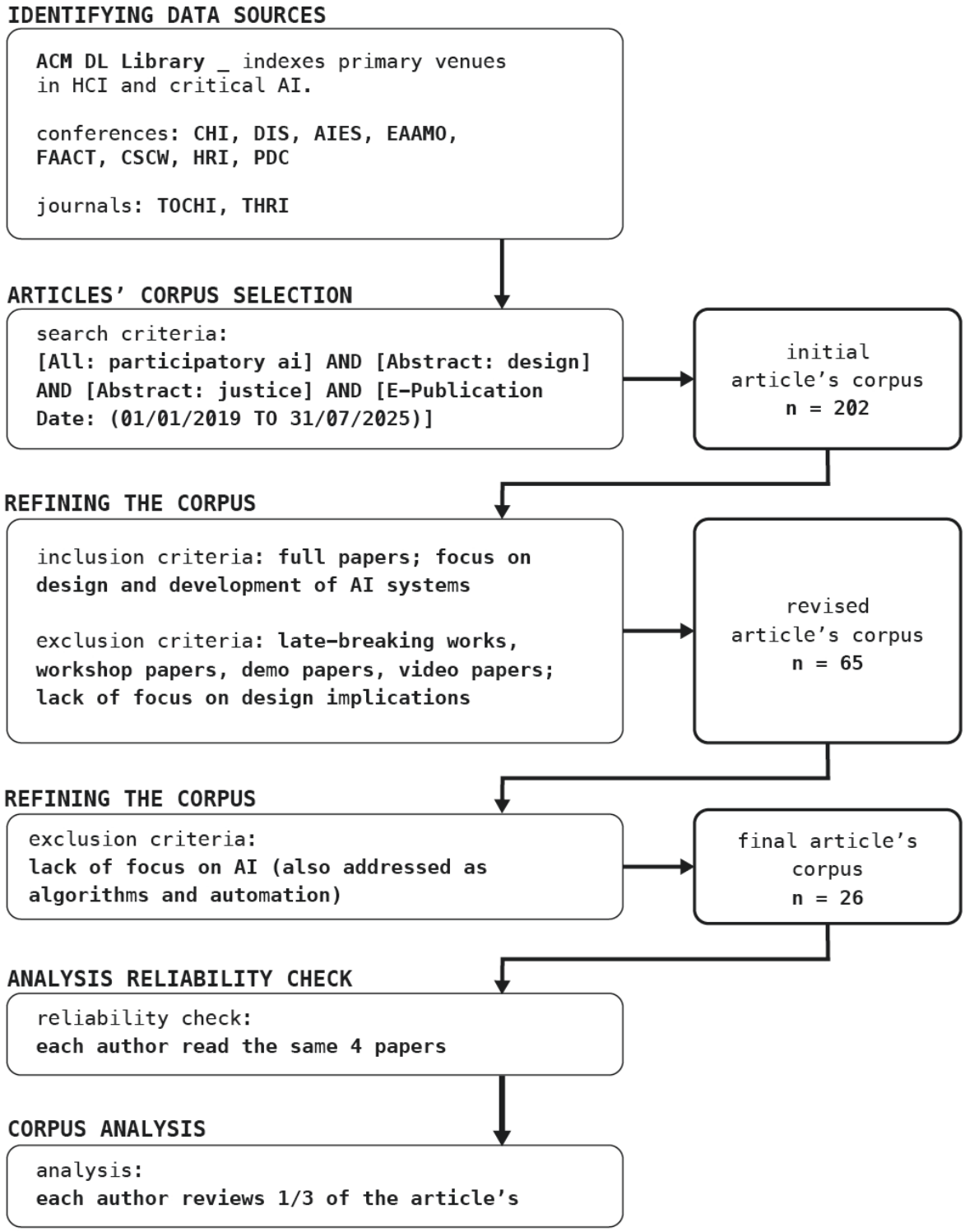


Figure 1: Overview of the scoping review process

(**TOCHI**), and ACM Transactions on Human-Robot Interaction Journal (**THRI**).

3.2.2 Search strategy. The search terms were designed to capture research explicitly engaging with participatory and justice-oriented approaches in AI:

```
All: participatory ai] AND [Abstract: design] AND
 [Abstract: justice] AND [E-Publication Date:
(01/01/2019 TO 31/07/2025)]
```

Articles were included if they contained the phrase '***participatory AI***' anywhere in the text, and '***design***' and '***justice***' in the abstract. The search covered publications from January 2019 to July 2025. We set January 2019 as the starting date for our corpus, as this corresponds to the period in which the EU High-Level Expert Group on AI presented the Ethics Guidelines for Trustworthy Artificial Intelligence [57], which marked a key milestone in the institutionalisation of participatory and ethical concerns in AI. The official presentation of the guidelines followed the publication of the first draft in December 2018, which received more than 500 comments through an open consultation. The initial search, conducted between May and July 2025, resulted in a corpus of 202 articles.

3.2.3 Exclusion and Inclusion Criteria. We limited the review to research articles as they are thoroughly peer-reviewed, excluding other formats such as demonstrations, late-breaking works, workshop proposals, student design competitions, and video presentations. In line with our disciplinary standpoint in design research and our commitment to a holistic understanding of participatory AI processes, we focused this investigation on participatory processes dedicated to the design and development of an AI system. We excluded studies that conducted participatory practices with the intent of understanding AI's impact on people and society, without providing clear insights and recommendations for designing AI systems. All three authors collaborated iteratively in the selection process. Titles and abstracts of the 202 articles were screened, with the corpus divided into three sub-corpora for efficiency. We checked whether the work actually addressed the theme of participatory design, social justice, and the design of AI systems. The absence of at least one of these themes represented an exclusion criterion. This phase resulted in a reduced corpus of **65 articles**. After independent annotations of the three authors, we further excluded articles that appeared not to be related to AI systems. The resulting corpus comprised **26 articles**, which served as the basis for subsequent analysis.

3.3 Overview of the Corpus and Data Analysis

Following corpus selection, we carried out a deductive thematic analysis [21, 86]. A deductive approach was chosen, rather than a reflexive thematic analysis, to enable systematic examination of how current scholarship incorporates existing critical theory and related designerly methodologies in participatory AI. This approach aligned with the overarching goal of the review, which was to map and assess how participation and commitments to social justice are conceptualised and operationalised within the field.

The theoretical background of our analysis is informed by feminist and decolonial critiques of AI [12, 22, 27, 47, 67, 69, 83], joining current efforts within the HCI community to articulate the limits of participatory AI better. In particular, we share with Delgado et al. [29] and Corbett et al. [26] a commitment to examining how power is (or is not) challenged and redistributed through participatory processes. Distinctively, however, our work aims to address various dimensions and implications of engaging with matters of social justice in participatory AI. Similar to Feffer et al. [41], we aim to provide a more holistic understanding of this practice space, extending beyond the focus on who is involved and how. For this, we specifically structured our analysis around the Design Justice Principles [33] (Table 1), which encourage reflection not only on the subjects and modes of participation, but also on the very intentions behind design interventions, the role of 'designers', and the extent to which alternative ways of knowing are genuinely recognized and valued [28, 44].

3.4 Overview of the Corpus

To contextualise the corpus and identify scholarly and geographic contexts relevant to power and epistemic dynamics [44, 70], we annotated each article by publication year, country of author affiliation, publication venue, and disciplinary background of the article. Publication year was recorded as reported in the article. The country of affiliation was derived from the institutional affiliation of the authors. Disciplinary background was coded using the **ACM Computing Classification System**.

3.5 Codebook Development

The codebook was developed deductively, building on critical theory and participatory design, as articulated in the ten Design Justice Principles [33]. The three authors collectively discussed and grouped these principles into guiding themes, which were then translated into analytical questions for annotating the corpus. Each theme was defined with reference to relevant principles, accompanied by a set of annotation dimensions. The codebook was iteratively refined through repeated annotation and discussion until consensus was reached. The final codebook comprised five thematic foci, each associated with specific Design Justice Principles:

- **Why**. *What was/should be the purpose/ambition of the design?* (relates to principles 1 and 9). This theme centres on the goal of the design practice and/or research. It helps to articulate whether the participatory practices are oriented to support, sustain, empower and heal communities in a non-extractive way with a clear emancipatory or liberatory orientation. The goal of the design is non-exploitative and helps connection rather than disconnection.
- **What**. *What is being designed? Is it a tool for the community? Is it AI? What kind of design work is the research trying to inform?* (relates to principles 8 and 10). This theme centres on what is being designed or produced through research, and the intended use of resulting knowledge or artefacts.
- **Who Involves**. *What was/should be the role of the designer?* (relates to principle 5) This theme centres on who initiates participation, the role of designer(s) or researcher(s), and their awareness about power dynamics, situatedness of the participatory practices and the type of partnership established with the participants.
- **Who is involved**. *Which community was or should be involved?* (relates to principle 2) This theme centres on which

Table 1: The *Design Justice Principles* [33] that guided the data analysis

Principle	Description
Principle 1	We use design to sustain, heal, and empower our communities, as well as to seek liberation from exploitative and oppressive systems
Principle 2	We centre the voices of those who are directly impacted by the outcomes of the design process
Principle 3	We prioritise design's impact on the community over the intentions of the designer
Principle 4	We view change as emergent from an accountable, accessible, and collaborative process, rather than as a point at the end of a process
Principle 5	We see the role of the designer as a facilitator rather than an expert
Principle 6	We believe that everyone is an expert based on their own lived experience, and that we all have unique and brilliant contributions to bring to a design process
Principle 7	We share design knowledge and tools with our communities
Principle 8	We work towards sustainable, community-led and -controlled outcomes
Principle 9	We work towards non-exploitative solutions that reconnect us to the earth and to each other
Principle 10	Before seeking new design solutions, we look for what is already working at the community level. We honour and uplift traditional, indigenous, and local knowledge and practices

communities or participants are (or should be) involved in the design process and how they are described.

- **How**. *How does or should the participation unfold?* (relates to principles 3, 4, 6 and 7) This theme centres on how participation is established and mediated by designed artefacts.

During iterative annotation to reach consensus over the corpus and codebook, we added an additional theme (reaching six in total) to account for the phase of the design process of the participatory practices. The resulting question and thematic description are as follows:

- **When**. *When in the AI lifecycle does the participation unfold or should unfold?* This theme takes into account the definition of Corbett et al. 2023 [26], considering the lifecycle process as comprising four interconnected steps, i.e., problem formulation, dataset development, model design and training, and deployment and monitoring.

3.6 Data analysis and reliability

To ensure the reliability of the themes and guiding questions, four randomly selected articles (14% of the final dataset) were independently screened, read, and annotated by all three authors. The resulting annotations were compared and discussed in detail to strengthen the reliability of the codebook. One article was subsequently excluded due to a lack of thematic relevance. We conducted several consensus discussion rounds and refined the questions after reading each of the three articles. Following the codebook refinement and consensus discussions, the full dataset was divided evenly among three authors, with each coding one-third of the corpus (9 articles each). The data analysis was performed by reading the abstracts, introduction, methods section, positionality statements (if available), discussions, and conclusions. When necessary, full texts were carefully read. On average, each article took about 25 minutes to review. The authors used the guiding questions and related definitions to annotate the articles and noted the instances of certain facets being present or not, or added answers to the guiding questions.

4 Findings

In this section, we first provide an overview of the pool of articles. Then, we detail the findings of the deductive thematic analysis organised per guiding question (i.e., why, what, who is involved, who involves, how and when).

4.1 Overview of The Corpus of Articles

4.1.1 Year of publication. The majority of articles included in the corpus were published in 2023 and 2024, with eight articles each [1, 30, 36, 42, 52, 61, 87, 97, 107]. Three articles were published in each of the years 2021 [46, 84], 2022 [37, 48, 95], and 2025 [68, 102, 105]. Only one article in the corpus was published in 2019 [112].

4.1.2 Country of affiliation. The articles in our corpus predominantly report research conducted by researchers based in WEIRD countries, particularly the United States which accounts for 17 of the 26 articles [1, 30, 36, 37, 42, 48, 52, 61, 63, 68, 84, 95, 97, 105, 106, 112, 113]. The United Kingdom (UK) follows with two articles [46, 64]. Other countries represented include Canada [87], Denmark [ref from the corpus], The Netherlands [8], and Sweden [60], each with one article in the corpus. Four articles resulted from international collaborations, namely one articles [53] with researchers based in the USA and Canada, one article [9] with researchers based in the USA and UK, one [107] between UK and Ireland; and a collaboration between UK, Bangladesh and USA [102] which is the only article in the corpus with a research institute based in the Global South.

4.1.3 Venue of publication. Most of the articles in our corpus were published at the CHI conference (nine out of 26 articles [46, 52, 53, 63, 87, 97, 105, 112, 113]). This is followed by the TOCHI journal with four articles [42, 68, 106, 107], and DIS, with another four articles [9, 30, 60, 102]. The AIES, EEAMO and FAccT conferences are represented in the corpus with each three articles [1, 8, 36, 37, 48, 61, 64, 84, 95].

Table 2: The *Design Justice Principles* [33] that guided the data analysis

Author(s)	Title	Venue	Year
A. Petterson, K. Cheng, and P. Chandra	Playing with Power Tools: Design Toolkits and the Framing of Equity	CHI	2023
A. A. Abdu, L. M. Chambers, D. K. Mulligan, and A. Z. Jacobs	Algorithmic Transparency and Participation through the Handoff Lens: Lessons Learned from the U.S. Census Bureau's Adoption of Differential Privacy	FAccT	2024
A. Thieme, M. Hanratty, M. Lyons, J. Palacios, R. F. Marques, C. Morrison, and G. Doherty	Designing Human-centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment	TOCHI	2023
Anna Kawakami, Amanda Coston, Haiyi Zhu, Hoda Heidari, and Kenneth Holstein	The Situate AI Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder, Early-stage Deliberations Around Public Sector AI Proposals	CHI	2024
A. Arzberger, S. Buijsman, M. L. Lupetti, A. Bozzon, and J. Yang	Nothing Comes Without Its World – Practical Challenges of Aligning LLMs to Situated Human Values through RLHF	AIES	2024
A. Z. Asha, S. Sultana, H. A. He, and E. Sharlin	Shotitwo First!": Unravelling Global South Women's Challenges in Public Transport to Inform Autonomous Vehicle Design	DIS	2024
B. Gansky and S.McDonald	CounterFAccTual: How FAccT Undermines Its Organizing Principles	FAccT	2022
B. A. Halperin, G. Hsieh, E. McElroy, J. Pierce, and D. K. Rosner	Probing a Community-Based Conversational Storytelling Agent to Document Digital Stories of Housing Insecurity	CHI	2023
C. D. Whitney, T. Naval, E. Quepons, S. Singh, S. R Rick, and L. Irani	HCI Tactics for Politics from Below: Meeting the Challenges of Smart Cities	CHI	2021
D. Wang, Q. Yang, A. Abdul, and B. Y. Lim	Designing Theory-Driven User-Centric Explainable AI	CHI	2019
E. Edenberg and A. Wood	Disambiguating Algorithmic Bias: From Neutrality to Justice	AIES	2023
I. G. Johnson and C. Crivellaro	Opening Research Commissioning To Civic Participation: Creating A Community Panel To Review The Social Impact of HCI Research Proposals	CHI	2021
J. Solyst, S. Xie, E. Yang, A. E.B. Stewart, M. Eslami, J. Hammer, and A. Ogan	"I Would Like to Design": Black Girls Analyzing and Ideating Fair and Accountable AI	CHI	2023
J. Taylor and A. Bruckman	Mitigating Epistemic Injustice: The Online Construction of a Bisexual Culture.	TOCHI	2024
J. Nee, G. Macfarlane Smith, A. Sheares, and I. Rustagi	Advancing social justice through linguistic justice: Strategies for building equity fluent NLP technology.	EEAMO	2021
J. Kay, A. Kasirzadeh and S. Mohamed	Epistemic injustice in generative AI.	AIES	2024
L. E. Forlano and M. K. Halpern	Speculative Histories, Just Futures: From Counterfactual Artifacts to Counterfactual Actions	TOCHI	2023
MD R. Haque, D. Saxena, K. Weathington, J. Chudzik, and S. Guha	Are We Asking the Right Questions?: Designing for Community Stakeholders' Interactions with AI in Policing	CHI	2024
M. Sloane, E. Moss, O. Awomolo, and L. Forlano	Participation Is not a Design Fix for Machine Learning	EEAMO	2022
P. Jääskeläinen, A. Holzapfel, and E. Eriksson	AI Art for Self-Interest or Common Good? Uncovering Value Tensions in Artists' Imaginaries of AI Technologies	DIS	2024
S. Sultana, H. M. Chowdhury, Z. Sultana, and N. Verdezoto	'Socheton': A Culturally Appropriate AI Tool to Support Reproductive Well-being	DIS	2025
T. R Dillahunt, A. J. Lu, and J. Velazquez	Eliciting Alternative Economic Futures with Working-Class Detroiters: Centering Afrofuturism in Speculative Design	DIS	2023
T. Tanksley, A. D. R. Smith, S. Sharma, and E. W. Huff Jr	Ethics is not neutral": Understanding Ethical and Responsible AI Design from the Lenses of Black Youth	CHI	2025
T. Jegede, M. K. Gerchick, A. S. Mathai, and A. Horowitz	Challenge Accepted? A Critique of the 2021 National Institute of Justice Recidivism Forecasting Challenge	EEAMO	2023
U. Ehsan, R. Singh, J. Metcalf, and M. Riedl	The Algorithmic Imprint	FAccT	2022
X. Koo, L. Scott, and A. Bruckman	Metrics and Macchiatos: Challenges for Service-Industry Wdustry Workers and the Need for Worker-Driven ICTs	TOCHI	2025

4.1.4 Disciplinary background of the article as per CSS Classification. There are four macro subsets of disciplinary backgrounds described by category and concepts of the computing field: (i) Computing Methodologies and Applied Computing, with four articles [36, 48, 84, 107]), (ii) Human-centred computing, with eighteen articles [9, 30, 37, 52, 60, 63, 68, 106, 113], (iii) Social and Professional topics, with three articles [1, 97, 105] and HCI design and evaluation methods, with one article [87]. Notably, one article does not report any CSS classification [8]. The Computing Methodologies and Applied Computing articles are further classified as focusing on a specific technical application (i.e., machine learning, natural language processing, law applied computing, 3 articles [48, 84, 107]) or they are related to computing philosophy and theory (1 article [36]). The Human-centred computing articles have various subclassifications, but they are predominantly classified as field and empirical studies and HCI theory. The Social and professional topics have various subclassifications, including education, government policy, and race and ethnicity. The HCI design and evaluation

methods article present in the corpus is the second one to make an explicit participatory design classification [87].

4.2 Deductive Thematic Analysis

4.2.1 ***Why****. What was/should be the purpose/ambition of the design?* The articles in our corpus reflect a wide range of purposes and ambitions in the participatory design of AI. Across this diversity, we identify three overarching purposes: 1) *Designing for justice*; 2) *Developing responsible AI*, and 3) *Empowering communities. Designing for justice* is devised in 12 out of 26 articles [1, 37, 42, 46, 52, 60, 64, 84, 95, 102, 106, 113] and manifest as a strong and explicit orientation towards tackling power structures, inequality and discrimination; and typically propose alternatives to the status quo through frameworks, methods or co-design practices. The articles emphasise celebrating and valuing the lived experience of communities and sharing agency in design processes. We find 9 out of 26 articles [8, 30, 48, 53, 61, 63, 87, 107, 112] committed to *Developing responsible AI* (e.g., co-developing an explainable diagnostic AI framework with clinicians [112] or co-designing a toolkit to support democratic decision making around public AI systems [63]), mostly through participatory engagements, i.e., a series of approaches to develop explainable, accountable, fair, transparent AI systems aligning with socio-ethical and legal values. Unlike the justice-oriented works, these articles do not always foreground lived experience but instead position communities or stakeholders as contributors to making AI more accountable. A minority of works (4 out of 26 articles [9, 68, 97, 105]), focus on *Empowering communities* directly engage with marginalized communities or underrepresented groups, creating opportunities for people to exercise control and agency in AI design processes. For example, Tanskey et al. [105] conducted participatory research with Black Youth, enabling them to discuss and speculate about the impact of racist AI technologies, while fostering an educational program where Black students imagine technology otherwise. Koo et al. [68] studied service workers' experiences with automated management systems, giving workers greater power over technological mediation in their workplaces.

4.2.2 ***Who Involves****. What was/should be the role of the designer?* Several articles offer theoretical or methodological frameworks with no direct mention of who involves. In most articles, researchers take the central role in PAI work, either as facilitators of processes, enablers of co-creation, or experts guiding user studies. However, the ways in which researchers enact this role vary significantly, ranging from maintaining clear boundaries between "researcher" and "participant" to cultivating more symmetrical partnerships where these roles blur. Four articles [9, 30, 52, 113] report work where researchers establish a truly symmetrical partnership, making the boundaries between researchers and participant communities less discernible. These cases blur the boundaries between "who involves" and "who is involved." For example, Whitney et al. [113] describe researchers serving on city coalitions and steering committees while community members contributed directly to research activities. Similarly, Halperin et al. [52] emphasise that their research was built upon "*pre-established relationships from sustained, ongoing community-engagement*" (p.7). Asha et al. [9], instead, focus on where teams of researchers and women from the Global South together make sense of the lived experience of navigating public transportation. In three articles [97, 105], researchers also act as teachers who involve their students in participatory activities that combine learning with critical engagement. One article [63] uniquely reports the involvement of professional facilitators from a private company, underscoring how facilitation roles can extend beyond academia.

Despite the centrality of the researcher's role in PAI, only eight out of 26 articles in the corpus explicitly report a positionality statement [9, 30, 60, 68, 97, 102, 105, 106]. These articles are primarily empirical and design studies, reporting how researchers' positionality impacts the research. For example, Taskley et al. [105] reflect that "*Our identities, scholarship, and positionalities shape the lens through which we analyse these systems, allowing us to critically engage with AI's role in exacerbating existing inequities and to offer frameworks that centre justice and equity*". Koo et al. [68] provide a detailed account of how relationships between researchers and service workers are deeply entangled with broader labour relations.

4.2.3 ***Who is involved****. Which community was or should be involved? OR who are the participants (factual or intended)?* The corpus reveals a wide diversity of communities, experts, and stakeholders involved in participatory AI projects. Several articles involve specific communities embedded in distinct social contexts. For example, Black girls critically reflecting on AI accountability [97], women in the Global South addressing safety in public transportation [9], creative practitioners across artistic fields [60], coalitions of community organisations in urban governance [113]. Other articles primarily involve professional experts in AI technology and various other fields in the participatory process. For instance, Wang et al. [112] worked with clinicians to co-design explainable AI, while Haque et al. [53] involved technical specialists and policy experts. These projects highlight the importance of domain expertise for shaping responsible and context-sensitive AI systems. Several articles with a more theoretical or methodological orientation do not involve participants directly but instead call for the engagement of broader communities in participatory AI practices. For example, Abdu et al. [1] and Sloane et al. [95] advocate for expanding the scope of participation to a broader array of "stakeholders", while Nee et al. [84] specifically call for the participation of diverse (unspecified) language communities in the design of natural language processing technologies.

4.2.4 ***What****. What is being designed? Is it a tool for the community? Is it AI? What kind of design work is the research trying to inform?* We observe the corpus dividing into three broad categories of outcomes: *Design Research*, *Theory and Methods*, and *User Study Research*. Seven articles [42, 46, 60, 97, 105, 112, 113] manifest a clear Design Research orientation. These engage directly with co-creation, speculative design and co-prototyping sessions to shape PAI processes and materialise alternatives to the status quo. These studies frequently use artefacts to support reflection and debate, such as mockups [112] and canvases [30, 60, 113]. Ten articles [1, 8, 36, 37, 48, 61, 64, 84, 87, 95] are oriented more towards producing *Theory and Methods*. These present critical essays, position articles and theoretical and methodological frameworks as critical lenses to address PAI. For example, Abdu et al. [1] present a framework to make visible how values are embedded within sociotechnical AI systems and to configure who participates in their

design. Nee et al. [84], instead, introduce a framework to help NLP researchers engage with matters of linguistic justice.

4.2.5 ***How****. How does or should the participation unfold?* In general, we observed that articles tend not to explicitly report on "how" participation unfolds; hence, we inferred it from the text and the methodology reported in the article. What emerges is a spectrum of approaches. The majority of articles [9, 36, 37, 46, 52, 53, 61, 63, 68, 102] seem oriented towards an informing relationship, that is, an exchange of information between the research community or participants about the design of AI systems. Several articles advocate for or try to implement a partnership [30, 42, 87, 105, 113] where power and agency are shared and there is space for reflexivity. The rest of the articles seem more inclined towards co-research relationships, that is co-understanding, co-creation and co-evaluation of technology without a formal attention to power and agency dynamics between researchers and non-researchers. We have observed a tester relation [107], i.e., involvement to test technology and consulting partnership [64, 106, 112], where non-researchers and non-designers are asked for input. Further, the majority of the articles recognises the value of lived experience and centre these in the design or sensemaking processes. Engagement with lived experience is seen as an immersion of reflexivity and a continuous productive confrontation [113], a key process for improving the design of AI systems sensitive to the life experience of marginalised groups [36].

Another interesting finding is how the majority of the articles [1, 8, 9, 42, 48, 52, 60, 63, 64, 84, 87, 95, 97, 102, 105, 112, 113] engages tools, toolkits, prototypes and artefacts to mediate participation. Co-design tools, activities and materials support participation as well as actual prototypes from hackathons. In a policy-oriented context, in particular, co-designed artefacts or pieces of technology become **boundary objects** [99] to, for example, mediate governmental adoption of AI systems. However, as noted by Abdu et al. [1] "*The mobilisation of any given boundary object is dependent not only upon the object itself, but also upon the motivation and orientation of those brokers that span and connect communities*". Speculative activities and speculative design, too. For example, Dillahunt et al. [30] use a speculative Afrofuturist workbook Building Utopias workbook, to probe participation and to encourage speculative and critical design thinking from an Afrofuturism lens in participatory AI processes. Even the materiality of AI is considered by authors a way to engage with the sense-making of participation [64]: material engagement with the AI mediate the experience "*AI can be an interactive tool for exploring one's own experiences. Image generation can be used to re-imagine and express oneself*".

Lastly, while 8 works [1, 36, 60, 61, 63, 64, 87, 107] do not explicitly articulate this, the majority of articles consider participation as key for the social change of our society. As discussed by Forlano and Harper [42]: "*As technology continues to play an essential, and even an existential role in the future of society, it is vital to continue to find ways of critically engaging the public with the ethical and political stakes around these choices*". Social change is often intended as a challenge to the status quo (as in [46]) and a moment of confrontation with the social implications of PAI actions, as the "*participation of all the community members might raise some derivative concerns, including resource constraints, an unfair amount of extra workload, stress, etc* [102].

4.2.6 ***When****. When in the AI lifecycle should the participation unfold or should unfold?* The majority of articles [1, 9, 30, 36, 42, 46, 52, 60, 61, 63, 64, 68, 84, 97, 102, 105, 106, 113] focus on the **problem definition** phase, while the rest of the corpus focuses on participation unfolding in the **deployment** and **monitoring phase** [37, 53, 107, 112], and the **complete lifecycle** [8, 48, 87, 95].

Research articles that highlight the importance of community involvement in the problem definition activities argue that focusing on this early phase of development is crucial for addressing socio-technical issues before actual technology development begins. Edenberg and Wood [36], for instance, call us to "*redirect approaches to addressing bias in generative AI at its early stages in ways that can more robustly meet the demands of justice*" (p.691). These works focused on the problem definition phase are primarily oriented to understanding people and communities in context and include many processes and methods such as surveying communities [1, 68, 102, 106], collective ideation and prototyping [46], the use of mock-ups [107] and speculative sessions [30, 42, 105].

In the deployment and monitoring phase, participation unfolds to understand possible opportunities for re-design as well as to understand the implications of deployment. Ehsan et al. [37] introduce the notion of "algorithmic imprinting" arguing that the consequences of algorithms extend further than the deployment period and even when an algorithm is eventually removed from a system and more attention should be given to situated fairness: "*Awareness and traceability of the contours of the imprint can facilitate improved active participation in algorithmic mediation. An imprint-aware design mindset treats stakeholders as active (as opposed to passive) participants in the design process*" (p. 1312 [37]).

Interestingly, of the articles focusing on the complete lifecycle or advocating for participation unfolding in the full process, a few provide interesting insights about current limitations related to "when participation unfolds". Gansky and McDonald [48], for instance, highlight how there is a tendency to "*focus on a single stage of development at a time, with the intention of affecting conditions across the entire life cycle*" and argue for more attention to be paid to the socio-technical and organisational context of AI deployment. In this regard, Ehsan et. al [37] call for extending the locus of analysis and looking at the junctions between use (lifecycle) and post-use (afterlife), for a more holistic assessment of algorithmic impact. In support of this, Petterson et al. [87] suggest that toolkits hold potential to support more holistic participation practices in extended AI lifecycles.

5 Discussion

Our analysis reveals that practising justice-oriented PAI work remains more an ambitious than a practical reality, as many of the articles present theoretical frameworks instead of empirical work. Nevertheless, there are several instances of how virtuous practices are instantiated and important reflections on our role and approaches are encouraged. In what follows, we provide further reflections on these methodological implications. We particularly

focus on the role of the PAI researcher/practitioner and the importance of building alliances with communities, along with a reflection on the important, yet not self-sufficient, role of artefacts in PAI processes.

5.1 Practising PAI, being an activist

What becomes clear by looking at virtuous examples doing the work of social justice in participatory AI is that the very identity of the PAI researchers and practitioners is a hybrid one. Powerful and holistic PAI work, in fact, is most often done by people who fit both the role of the researcher/practitioner and the one of participants in the research (or at least a member of the community that the research looks at). For example, Koo and colleagues' work [68] about the pressure of workplace technologies on service-industry workers stems from the personal experience of the second author working for two years as a barista at Starbucks and the first author's experience as a member of the United Campus Workers union. The engagement of participants in this study heavily and openly relies on the personal networks of the authors. In the same vein, the work by Whitney and colleagues [113] on understanding and influencing the design and implementation of smart streetlight infrastructures in the city of San Diego is strongly motivated by the last author being a coalition steering committee member. Relatedly, the work by Asha and colleagues [9] on the potential of autonomous vehicle technologies for women in the Global South is motivated by the daily personal mobility experience and struggles of the first two authors, who were born and raised in Bangladesh, and identify as women. Personal experience and interests of the PAI researchers/practitioners, then, represents not only a motivator for engaging with specific communities, but also a facilitator, as this allows for leveraging networks, eases communication, and may reduce hostility and mistrust.

When the "personal" is at stake, **PAI research becomes a form of activism**. This aligns with the heritage of design tradition and its commitment to socio-political matters. As Bieling [16] articulates in their book "Design (&) Activism: Perspectives on Design as Activism and Activism as Design", over the last decades, the increasing vigour, diversity and number of popular and social movements has encouraged an extensive use of designerly methods and tools in support of individuals and communities to help expressing themselves and having a voice in matters of public concerns. Design activism clearly has a strong tie to political activism, but also materializes in distinct acts: instead of boycotts, strikes, protests, or demonstrations, design activism "*lends its power of resistance by intervening in people's lives*" [75]––it focuses on developing artifacts that exist in real time and space, and are situated within everyday contexts and processes of social and economic life [62]. PAI as activism blends these two perspectives as it both instantiates forms of public aggregation and contestation, as well as makes use of artefacts to contest dominant socio-technical paradigms and facilitate the collective shaping of socially desirable alternatives. Design workshops and hackathons become sites where PAI researchers/practitioners join forces with community organisations and members to voice dissent and envision alternative socio-technical paradigms [30, 60, 63, 113], in which artefacts and mockups are designed and used to support dialogue [1, 52]–to function as *AI boundary objects*.

This hybrid identity of the PAI researcher/practitioner is a key enabler of meaningful engagement with communities. Still, it is also a **hard-to-maintain** and an **impossible-to-prescribe** dimension of these practices. As Harrington and colleagues [54] argue, in fact, equity-driven and community-based participatory design practices are well-accepted both inside and outside of academia, yet giving them continuity is hard, especially after fundings end. A different academic culture where alternative forms of impact are acknowledged is needed [54], but also the toll that these practices impose on individuals, both in terms of time [17] and emotional labour [54], must be understood. Once again, design scholarship underscores that there are limits to the affordability of critical practice, as designers (and by extension PAI researchers/practitioners) also have to work under capitalist conditions [76]. It is important to acknowledge that one can act from a critical standpoint, in radical and activist terms, at one precise moment in their career, but not later, because circumstances can change, both in the personal sphere and the disciplinary discourse [72]. But, *what happens when a PAI researcher/practitioner engagement with a community ends?*

If we are to claim the merits of justice in our work, it is also essential to account for the effects that our interventions have on communities and contexts, beyond the exact period of our intervention. We should avoid what we can metaphorically name '*pride parade*' engagements– acts of performative allyship that is showcased only at the time of an event [111]– and account for the consequences of us both entering and exiting a community. In 'traditional' design processes, the value and impact of a participatory design intervention can be assessed in the way a product, service or system that has been collectively developed, gets actually adopted and used [20]. When it comes to practising PAI with the ambition to fight for social justice, the very engagements, the gatherings, and the resulting forms of **contestation and resistance** constitute the value of the work [26, 49, 108]. But, *how can this be sustained over time?* Useful in this regard is to rethink once again about the relationship between who leads PAI work and who participates. Instead of approaching these practices as individuals or teams of sole academics, **we better build large alliances with diverse actors** [96]. Dillahunt and colleagues [30], for instance, report on the collective exploration of alternative economic futures with Black and Brown working-class Detroiters. The team in their research is uniquely composed of academic researchers, community leaders and activists. This alliance required an extensive investment of effort and time to define shared goals and ways of working together, but holds the potential to provide continuity to the cause, as it doesn't rely on the availability and individual conditions of single subjects. For alliances to work, however, it is of utmost importance to understand the organisational dynamics of collectives that are "*shaped and reshaped in changing networks, co-operations, more or less temporary organisational synergies*" [96] and to be open to **rethinking the identity of PAI researchers/practitioners from individuals to alliances**.

5.2 Practising PAI, crafting and using AI boundary objects

Central to many of the works reviewed here is an effort to provide the field with theoretical instruments to structure PAI work and/or

artefacts for facilitating engagement of participants and their deliberation. In both cases, we argue, the work can be framed as a practice of **crafting and using AI boundary objects**. This framing builds on the popular social science notion of *boundary objects*: artefacts that are robust enough to maintain a common identity across different social or disciplinary worlds, yet flexible enough to be interpreted locally by each group [99]. These are things that can be abstract or physical and "exist in the liminal spaces between adjacent communities of people and have the capacity to traverse perceptual and practical differences among these and facilitate cooperation by fostering mutual understanding" [58]. Because of this focus on the capacity to speak to different communities of practice and knowledge, the notion of boundary objects has established its role in innovation processes to enhance the capacity of an idea, theory, or practice to translate across culturally defined boundaries [43]. Within the space of participatory AI committed to social justice, boundary objects are potentially of crucial importance for their capacity to help understand, negotiate and break epistemic divides between researchers/practitioners and participants. AI boundary objects can help us challenge the "*AI's epistemological apartheid*" ([77] page 110) and the mirage of objectivity and scientism, creating bridges to alternative ways of knowing that are rooted in the lived experience of people who are marginalised or minoritised ([77] page 105).

A quick search of academic articles on AI boundary objects outputs works mentioning machine learning explanations [11], structural causal models [56], and onboarding materials [23] as boundary objects, just to name a few. This unveils a practical and technical view on AI boundary objects, where epistemic divides are only partially understood and addressed. Within our corpus, we find a few examples that align with this approach. Thieme and colleagues [107], for example, made use of mockups and their design rationale as a way to engage with experts to discuss clinically relevant AI applications, and the implications of design for AI applications in sensitive use contexts. Many of the reviewed works, however, make use of AI boundary *objects as artefacts materialising AI possibilities and to facilitate engagement in PAI work*. These do not necessarily engage with technical matters and value lived experiences over expert knowledge. In some cases, the AI boundary objects are themselves objects of collaborative design and become sites where the boundaries between the PAI researchers/practitioners and participants get blurred. Jääskeläinen and colleagues [60] structured their participatory activities as a blend of collaborative ideation session, during which each participant was invited to imagine an AI tool for their creative practice, and a collective discussion aimed at raising critical reflections on the imagined possibilities. Here the PAI researchers/practitioners did come with pre-made artefacts, i.e., *Speculative Sketching Templates*, but the very role of the AI boundary object was performed by the ideas represented on those templates. In a similar vein, Dillahunt and colleagues [30] provided participants with the *Building Utopias* workbooks and a card deck. These materials introduced new design and technology terms and served as a way for participants to both document their thoughts and collectively discuss. The work of Whitney and colleagues [113] is similar: they collaboratively designed artefacts with participants, during what they called a *Slightly Dystopian hackathon*. The work focused on developing speculative mockups of possible applications building on the APIs of smart streetlights in the city of San Diego. Distinctively, however, their work emphasises how the mockups were not much valued for the ideas they were embedding, but rather for the capacity of these to help the coalition working on the initiative to problematize a 'broken' public use of data and engage the community around the problem.

We also find examples of frameworks and theories that can help us structure PAI work and that can potentially be used by teams to reflect on their PAI study designs. Many of these, such as the ones by Sloane and colleagues [95] and Jegede and colleagues [61], propose frameworks for better understanding participation in AI development and deployment processes, and potential issues. In a similar vein, Abdu and colleagues [1] propose the concept of *handoff* as a critical tool for untangling the daunting knot of actors, components, modes, functions, and values that should be accounted for in collective processes of deliberation focused on algorithmic innovation. What this type of work underscores is the possibly dual nature of AI boundary objects. While on the one hand AI boundary objects can be used to ease technology and innovation adoption, on the other, these can also have the contrary effect, that is, to inhibit the uptake of a certain technology [43]. **AI boundary objects can function as a form of '*adversarial design*'** [31]: by manifesting controversial aspects of a technology, it opens up space for contestation and dissent–agoras for constructive agonism. As such, AI boundary objects hold great potential for practising PAI work that is committed to social justice: not only to achieve a more balanced distribution of knowledge between different subjects in participatory settings, but also to allow for agonistic feelings and opinions to be expressed.

Making AI boundary objects 'work', however, it's all but trivial. As Abdu and colleagues [1] articulate by reflecting on their study of the U.S. Census Bureau's adoption of differential privacy [2], the Bureau employed several conceptual boundary objects (i.e., interactive notebooks, webinars, handbooks, and videos) that resulted to be only partially successful. According to the authors, the causes of such limited success lies in an insufficient comprehension of the diverse ecosystem of stakeholders and a lack of connection with trusted local experts that could have carried the boundary objects across community divides. What the authors further underscore is that "boundary objects cannot travel alone" and it is crucial for the critical AI community to "*think about the expertise needed to shepherd and use such boundary objects effectively in order to broker meaningful trust and participation*" [1].

The craft and use of AI boundary objects in participatory AI, then, holds great potential for engaging with matters of social justice. However, it also comes with a multitude of disciplinary challenges. As PAI researchers/practitioners, we should ask what constitutes a good AI boundary object and what type of artefacts can function better to address specific aspects and implications of AI implementations. Along with such a focus on artefacts, however, **we should also instantiate models of engagement in which the craft and use of AI boundary objects is not a sole appannage of PAI researchers/practitioners, but rather an instrument of the broader PAI coalitions** we mentioned in the previous paragraphs. We should ultimately investigate, collectively, what are the cultural and contextual conditions that an AI boundary object should fit in and how to better speak to these through things.

6 Conclusions

This work draws a picture of and reflects upon the current landscape of PAI research and practice in HCI, with a particular focus on matters of social justice and whether these are addressed through design, and in what forms. Grounding of feminist and decolonial AI critiques, and guided by design justice principles, we highlight current best practices, identify opportunities for future methodological expansions, and take stock of existing challenges.

Our scoping review of 26 articles from HCI venues finds that:

(1) The most common drivers (why) for justice-oriented PAI work are the ambition to develop more *responsible AI* systems on the one hand, and the commitment to *empower communities* to stir AI innovation processes, including the possibility of actions for dismantling power structures, inequality and discrimination as a way to *design for justice*;
(2) The object of design (what) in these practices is often intangible and conceptual, rather than a physical thing. Most articles report on the development of theoretical or methodological frameworks and/or user studies that help us move in the space of PAI work;
(3) There generally is an explicit commitment to involve communities, yet the power dynamics involved in these engagements, including the position of the researchers/practitioners (who involves), are rarely addressed;
(4) Participants (who is involved) are generally field experts, stakeholders of some sort, and community members, sometimes with an explicit focus on marginalised communities. Virtuous cases engage deeply with community-led organisations, but these remain a minority;
(5) The majority of works engage with tools, toolkits, mockups and other artefacts to facilitate participation, yet these works tend to focus more on the procedural aspects of doing PAI work (how);
(6) The majority of works focus on the early phases of the AI development cycle (when);

We further reflect critically on these findings and observe that there is a necessary methodological broadening that comes when practising PAI work, which concerns the blurry boundaries between 'who involves' and 'who is involved'. Our observation, grounded on a few virtuous cases (e.g., [30, 96]), is that instead of structuring PAI work as an endeavour that is exclusively led by the researchers/practitioners, larger and long-term alliances should be built, especially with community-led initiatives and activist groups. This is not only essential for facilitating community engagement, but also for ensuring continuity and impact of the PAI initiatives beyond the times and resources that are available to the individual researchers/practitioners. And along with encouraging a reflection on our very role as design and HCI researchers, we also suggest rethinking once again the role that artefacts play within PAI processes. Our observation here is that artefacts hold enormous potential for facilitating justice-oriented PAI work, especially if used as *AI boundary objects*. As such, design and HCI researchers/practitioners have an important role to play. Once again, however, we must carefully investigate, collectively, the cultural and contextual conditions that an AI boundary object should fit in, and how PAI alliances can appropriate and make use of them.

To conclude, this work contributes to critical AI and more broadly to HCI and design, scholarship concerned with matters of social justice in participatory processes. To our knowledge, this is the first scoping review specifically focusing on PAI research in HCI explicitly analysed through the lens of social justice, and design justice more specifically. The overview we provide on the current state of the art of justice-oriented PAI work can help design, HCI and AI researchers/practitioners to navigate the field. In particular, it can serve as a reference point for reflecting on our position and standing, and for critically scrutinising the procedures we set up and questioning whether we are reinforcing power imbalances and non-equitable forms of participation.

Finally, the aim of this work is also to call our community once again to understand its political standing–to acknowledge the **deeply political nature of PAI work**. We also want to underscore the message, however, that taking a stand on these public matters does not necessarily mean to only be critical and 'against'. We would rather encourage our peers to **get inspired and celebrate the virtuous work that the community is producing, and to approach this research and practice field as a space of connection, care and collective growth**.

6.1 Limitations

We also acknowledge that our scoping review has some limitations. First of all, participatory design is an "in-situ", reflexive and holistic practice. The complexity of participatory practices often cannot be captured fully by published articles only "because the implementation of AI involves navigating complex issues across multiple organisational and political arenas and scales, including misaligned stakeholder expectations, concerns about data quality, and ethical dilemmas."([96] p.186). In addition, the choice to use the design justice principles as an analytical focus is a specific way to look at justice-oriented PAI. A combination of this with other frameworks can be helpful for gaining an even more complete picture of PAI and design justice work. Moreover, as PAI is an emergent field in HCI and AI, our choice of focusing on full papers might carry the consequence of valuable work being done, but only being published in short papers and late-breaking reports. Last, there is a plethora of important justice-oriented work, including PAI work, being done by community-led organisations and activist groups that do not end up being published in academic venues. As researchers committed to matters of social justice, we should build new approaches for learning from this body of non-academic work.

Acknowledgments

Thank you Gianna for the panna.

References

[1] Amina A Abdu, Lauren M Chambers, Deirdre K Mulligan, and Abigail Z Jacobs. 2024. Algorithmic transparency and participation through the handoff lens: Lessons learned from the US Census Bureau's adoption of differential privacy. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1150–1162.
[2] John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2867–2867.
[3] William Agnew, A Stevie Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R McKee. 2024. The illusion of

artificial inclusion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–12.
[4] William Agnew, Kevin R McKee, I Gabriel, J Kay, W Isaac, AS Bergman, S El-Sayed, and S Mohamed. 2023. Technologies of resistance to AI. *Equity and Access in Algorithms, Mechanisms, and Optimization* (2023), 1–13.
[5] Evgeni Aizenberg and Jeroen Van Den Hoven. 2020. Designing for human rights in AI. *Big Data & Society* 7, 2 (2020), 2053951720949566.
[6] Hilary Arksey and Lisa O'malley. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology* 8, 1 (2005), 19–32.
[7] Sherry R Arnstein. 1969. A ladder of citizen participation. *Journal of the American Institute of planners* 35, 4 (1969), 216–224.
[8] Anne Arzberger, Stefan Buijsman, Maria Luce Lupetti, Alessandro Bozzon, and Jie Yang. 2024. Nothing comes without its world–practical challenges of aligning llms to situated human values through RLHF. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 61–73.
[9] Ashratuz Zavin Asha, Sharifa Sultana, Helen Ai He, and Ehud Sharlin. 2024. "Shotitwo First!": Unraveling Global South Women's Challenges in Public Transport to Inform Autonomous Vehicle Design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 3193–3209.
[10] Jacqui Ayling and Adriane Chapman. 2022. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* 2, 3 (2022), 405–429.
[11] Amid Ayobi, Katarzyna Stawarz, Dmitri Katz, Paul Marshall, Taku Yamagata, Raúl Santos-Rodríguez, Peter Flach, and Aisling Ann O'Kane. 2021. Machine learning explanations as boundary objects: how AI researchers explain and non-experts perceive machine learning. (2021).
[12] Amir Baradaran. 2024. Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. *Ai & Society* 39, 1 (2024), 7–19.
[13] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 675–684.
[14] Zack Beauchamp. 2024. Why the far right is surging all over the world. https://www.vox.com/politics/361136/far-right-authoritarianism-germany-reactionary-spirit
[15] Eevi E Beck. 2002. P for political: Participation is not enough. *Scandinavian Journal of Information Systems* 14, 1 (2002), 1.
[16] Tom Bieling. 2019. *Design (&) activism: Perspectives on design as activism and activism as design.* Mimesis.
[17] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.
[18] Luigi Bobbio. 2019. Designing effective public participation. *Policy and Society* 38, 1 (2019), 41–57.
[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*. 425–436.
[20] Tone Bratteteig and Ina Wagner. 2016. What is a participatory design result?. In *Proceedings of the 14th Participatory Design Conference: Full papers-Volume 1*. 141–150.
[21] Virginia Braun and Victoria Clarke. 2022. Conceptual and design thinking for thematic analysis. *Qualitative psychology* 9, 1 (2022), 3.
[22] Jude Browne, Stephen Cave, Eleanor Drage, and Kerry McInerney. 2023. *Feminist AI: Critical perspectives on algorithms, data, and intelligent machines.* Oxford University Press.
[23] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2021. Onboarding materials as cross-functional boundary objects for developing ai assistants. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.
[24] Luciano Cavalcante Siebert, Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink, Elisa Giaccardi, Geert-Jan Houben, Catholijn M Jonker, et al. 2023. Meaningful human control: actionable properties for AI system development. *AI and Ethics* 3, 1 (2023), 241–255.
[25] Patricia Hill Collins. 2022. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment.* routledge.
[26] Eric Corbett, Remi Denton, and Sheena Erete. 2023. Power and public participation in ai. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–13.
[27] Sasha Costanza-Chock. 2018. Design justice: Towards an intersectional feminist framework for design theory and practice. In *Proceedings of the DRS International Conference 2018, 25-28 June, Limerick, Ireland*. 529–540.
[28] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.* The MIT Press.
[29] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–23.
[30] Tawanna R Dillahunt, Alex Jiahong Lu, and Joanna Velazquez. 2023. Eliciting alternative economic futures with working-class detroiters: Centering afrofuturism in speculative design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 957–977.
[31] Carl DiSalvo. 2015. *Adversarial design.* Mit Press.
[32] C DiSalvo, A Clement, and V Pipek. 2013. Communities: Participatory Design for, with and by Communities,[w:] Routledge International Handbook of Participatory Design, red. J. Simonsen, T. Robertson.
[33] DJN. 2018. Design Justice Network Principles. https://designjustice.org/read-the-principles
[34] DJN. 2025. Design Justice Network. https://designjustice.org/about-us
[35] Sibel Duz. 2025. Gaza as a testing ground: Israel's AI warfare. https://www.setav.org/en/gaza-as-a-testing-ground-israels-ai-warfare
[36] 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*. 691–704.
[37] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The algorithmic imprint. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1305–1317.
[38] Salma Elsayed-Ali, Elizabeth Bonsignore, and Joel Chan. 2023. Exploring challenges to inclusion in participatory design from the perspectives of global north practitioners. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–25.
[39] Motahhare Eslami, Sarah Fox, Hong Shen, Bobbie Fan, Yu-Ru Lin, Rosta Farzan, and Beth Schwanke. 2025. From Margins to the Table: Charting the Potential for Public Participatory Governance of Algorithmic Decision Making. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 2657–2670.
[40] Gregory Falco. 2019. Participatory AI: Reducing AI bias and developing socially responsible AI in smart cities. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, 154–158.
[41] Michael Feffer, Michael Skirpan, Zachary Lipton, and Hoda Heidari. 2023. From preference elicitation to participatory ML: A critical survey & guidelines for future research. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 38–48.
[42] Laura E Forlano and Megan K Halpern. 2023. Speculative histories, just futures: From counterfactual artifacts to counterfactual actions. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–37.
[43] Nick J Fox. 2011. Boundary objects, social meanings and the success of new technologies. *Sociology* 45, 1 (2011), 70–85.
[44] Sarah Fox, Mariam Asad, Katherine Lo, Jill P Dimond, Lynn S Dombrowski, and Shaowen Bardzell. 2016. Exploring social justice, design, and HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 3293–3300.
[45] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance* 77, 1 (2022), 5–47.
[46] Ian G Johnson and Clara Crivellaro. 2021. Opening research commissioning to civic participation: creating a community panel to review the social impact of HCI research proposals. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.
[47] Iason Gabriel. 2022. Toward a theory of justice for artificial intelligence. *Daedalus* 151, 2 (2022), 218–231.
[48] Ben Gansky and Sean McDonald. 2022. CounterFAccTual: How FAccT undermines its organizing principles. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1982–1992.
[49] Aakash Gautam. 2024. Reconfiguring participatory design to resist ai realism. In *Proceedings of the Participatory Design Conference 2024: Exploratory Papers and Workshops-Volume 2*. 31–36.
[50] Mathieu Edouard Samborska Veronika Giattino, Charlie and Max Roser. 2023. Data Page: Number of large-scale AI systems released per year. https://archive.ourworldindata.org/20250813-134018/grapher/number-of-large-scale-ai-systems-released-per-year.html
[51] Pauline Gourlet, Donato Ricci, and Maxime Crépel. 2024. Reclaiming artificial intelligence accounts: A plea for a participatory turn in artificial intelligence inquiries. *Big Data & Society* 11, 2 (2024), 20539517241248093.
[52] Brett A Halperin, Gary Hsieh, Erin McElroy, James Pierce, and Daniela K Rosner. 2023. Probing a community-based conversational storytelling agent to document digital stories of housing insecurity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[53] MD Romael Haque, Devansh Saxena, Katy Weathington, Joseph Chudzik, and Shion Guha. 2024. Are we asking the right questions?: Designing for community stakeholders' interactions with ai in policing. In *Proceedings of the 2024 CHI*

Conference on Human Factors in Computing Systems. 1–20.
[54] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–25.
[55] Steve Harrison, Phoebe Sengers, and Deborah Tatar. 2011. Making epistemological trouble: Third-paradigm HCI as successor science. *Interacting with computers* 23, 5 (2011), 385–392.
[56] Hans-Martin Heyn and Eric Knauss. 2022. Structural causal models as boundary objects in AI system development. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. 43–45.
[57] AI Hleg. 2019. Ethics guidelines for trustworthy AI. *B-1049 Brussels* (2019).
[58] Isto Huvila, Theresa Dirndorfer Anderson, Eva Hourihan Jansen, Pam McKenzie, and Adam Worrall. 2017. Boundary objects in information science. *Journal of the Association for Information Science and Technology* 68, 8 (2017), 1807–1822.
[59] IxDF. 2023. Participatory Design. https://www.interaction-design.org/literature/topics/participatory-design
[60] Petra Jääskeläinen, André Holzapfel, and Emelie Eriksson. 2024. AI Art for Self-Interest or Common Good? Uncovering Value Tensions in Artists' Imaginaries of AI Technologies. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2897–2910.
[61] Tobi Jegede, Marissa Kumar Gerchick, Amreeta S Mathai, and Aaron Horowitz. 2023. Challenge accepted? A critique of the 2021 National Institute of Justice recidivism forecasting challenge. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–17.
[62] Guy Julier. 2013. From design culture to design activism. *Design and Culture* 5, 2 (2013), 215–236.
[63] Anna Kawakami, Amanda Coston, Haiyi Zhu, Hoda Heidari, and Kenneth Holstein. 2024. The situate AI Guidebook: Co-designing a toolkit to support multi-stakeholder, early-stage deliberations around public sector AI proposals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.
[64] Jackie Kay, Atoosa Kasirzadeh, and Shakir Mohamed. 2024. Epistemic injustice in generative AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 684–697.
[65] Finn Kensing and Jeanette Blomberg. 1998. Participatory design: Issues and concerns. *Computer supported cooperative work (CSCW)* 7, 3 (1998), 167–185.
[66] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A Hale. 2023. Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. *arXiv preprint arXiv:2303.05453* (2023).
[67] Lauren Klein and Catherine D'Ignazio. 2024. Data feminism for AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 100–112.
[68] Xander Koo, Lucy Scott, and Amy Bruckman. 2025. Metrics and Macchiatos: Challenges for Service-Industry Wdustry Workers and the Need for Worker-Driven ICTs. *ACM Transactions on Computer-Human Interaction* (2025).
[69] Tanja Kubes. 2025. A Critique of Human-Centred AI: A Plea for a Feminist AI Framework (FAIF). *AI & SOCIETY* (2025), 1–12.
[70] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How weird is CHI?. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–14.
[71] Thomas Lodato and Carl DiSalvo. 2018. Institutional constraints: the forms and limits of participatory design in the public realm. In *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1*. 1–12.
[72] Maria Luce Lupetti. 2022. Understanding critical design tensions with progression through historical practices. *Diid—Disegno Industriale Industrial Design* 77 (2022), 12.
[73] Maria Luce Lupetti and Dave Murray-Rust. 2024. (un) making ai magic: a design taxonomy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
[74] Maria Isabel Magaña and Katie Shilton. 2025. Frameworks, Methods and Shared Tasks: Connecting Participatory AI to Trustworthy AI Through a Systematic Review of Global Projects. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 2166–2179.
[75] Thomas Markussen. 2013. The disruptive aesthetics of design activism: enacting design between art and politics. *Design Issues* 29, 1 (2013), 38–50.
[76] Afonso Matos. 2022. *Who Can Afford to be Critical?: An Inquiry Into what We Can't Do Alone, as Designers, and Into what We Might be Able to Do Together, as People*. Set Margins'.
[77] Dan McQuillan. 2022. *Resisting AI: An anti-fascist approach to artificial intelligence*. Policy Press.
[78] Naja Holten Møller, Claus Bossen, Kathleen H Pine, Trine Rask Nielsen, and Gina Neff. 2020. Who does the work of data? *Interactions* 27, 3 (2020), 52–55.
[79] Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Commun. ACM* 36, 6 (1993), 24–28.
[80] N/A. 2024. With Global Crises Increasing Levels of Poverty, Unprecedented Investments Will Be Needed to Reverse Trend, Speakers Tell Second Committee. https://press.un.org/en/2024/gaef3607.doc.htm
[81] N/A. 2025. Facts about the nature crisis. https://www.unep.org/facts-about-nature-crisis
[82] N/A. 2025. Tech giants see emissions surge 150 percent in 3 years amid AI boom: UN. https://www.aljazeera.com/news/2025/6/6/tech-giants-see-emissions-surge-150-percent-in-3-years-amid-ai-boom-un
[83] Angella Ndaka, Philippe JC Lassou, Konan Anderson Seny Kan, and Samuel Fosso-Wamba. 2024. Toward response-able AI: A decolonial perspective to AI-enabled accounting systems in Africa. *Critical Perspectives on Accounting* 99 (2024), 102736.
[84] Julia Nee, Genevieve Macfarlane Smith, Alicia Sheares, and Ishita Rustagi. 2021. Advancing social justice through linguistic justice: Strategies for building equity fluent NLP technology. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–9.
[85] Ma Michelle Nicoletti, Leonardo and Dina Bass. 2025. AI is draining water from areas that need it most. https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/
[86] Noel Pearse. 2019. An illustration of deductive analysis in qualitative research. In *18th European conference on research methodology for business and management studies*. 264.
[87] Adrian Petterson, Keith Cheng, and Priyank Chandra. 2023. Playing with power tools: Design toolkits and the framing of equity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–24.
[88] Jennifer Pierre, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. Getting Ourselves Together: Data-centered participatory design research & epistemic burden. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.
[89] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, et al. 2023. Queer in AI: A case study in community-led participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1882–1895.
[90] Ben Quinn and Dan Milmo. 2024. How the far right is weaponising AI-generated content in Europe. https://www.theguardian.com/technology/2024/nov/26/far-right-weaponising-ai-generated-content-europe
[91] María Méndez Rocasolano. 2022. Human rights, big data and artificial intelligence: Elements of a complex algorithm. In *Security and Defence: Ethical and Legal Challenges in the Face of Current Conflicts*. Springer, 93–102.
[92] Elizabeth B-N Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *Co-design* 4, 1 (2008), 5–18.
[93] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew BA McDermott, Irene Y Chen, and Marzyeh Ghassemi. 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine* 27, 12 (2021), 2176–2182.
[94] Mona Sloane. 2024. Controversies, contradiction, and "participation" in AI. *Big Data & Society* 11, 1 (2024), 20539517241235862.
[95] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–6.
[96] Rachel Charlotte Smith, Liesbeth Huybrechts, Jesper Simonsen, and Daria Loi. 2025. Contemporary Participatory Design: Research Agendas for Societal Crisis. In *Proceedings of the sixth decennial Aarhus conference: Computing X Crisis*. 182–201.
[97] Jaemarie Solyst, Shixian Xie, Ellia Yang, Angela EB Stewart, Motahhare Eslami, Jessica Hammer, and Amy Ogan. 2023. "I would like to design": Black girls analyzing and ideating fair and accountable AI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
[98] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication* 52, 2 (2005), 163–174.
[99] Susan Leigh Star and James R Griesemer. 1989. Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science* 19, 3 (1989), 387–420.
[100] Paul Stares. 2025. Conflicts to Watch in 2025. https://www.cfr.org/report/conflicts-watch-2025
[101] Marc Steen. 2013. Co-design as a process of joint inquiry and imagination. *Design issues* 29, 2 (2013), 16–28.
[102] Sharifa Sultana, Hafsah Mahzabin Chowdhury, Zinnat Sultana, and Nervo Verdezoto. 2025. 'Socheton': A Culturally Appropriate AI Tool to Support Reproductive Well-being. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 3098–3116.
[103] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1609–1621.
[104] Mariarosaria Taddeo and Luciano Floridi. 2018. How AI can be a force for good. *Science* 361, 6404 (2018), 751–752.
[105] Tiera Tanksley, Angela DR Smith, Saloni Sharma, and Earl W Huff Jr. 2025. " Ethics is not neutral": Understanding Ethical and Responsible AI Design from

the Lenses of Black Youth. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[106] Jordan Taylor and Amy Bruckman. 2024. Mitigating epistemic injustice: The online construction of a bisexual culture. *ACM Transactions on Computer-Human Interaction* 31, 4 (2024), 1–34.

[107] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.

[108] Johannes Thumfart. 2024. The democratic offset: Contestation, deliberation, and participation regarding military applications of AI. *AI and Ethics* 4, 2 (2024), 511–526.

[109] Loreben Tuquero. 2024. Did artificial intelligence shape the 2024 US election? https://www.aljazeera.com/news/2024/12/25/did-artificial-intelligence-shape-the-2024-us-election

[110] Barbara-Chiara Ubaldi and Ricardo Zapata. 2024. GOVERNING WITH ARTIFICIAL INTELLIGENCE: ARE GOVERNMENTS READY? https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/governing-with-artificial-intelligence_f0e316f5/26324bc2-en.pdf

[111] Bao-Thi Van Cong, Séamus A Power, and Thomas A Morton. 2025. Solidarity with whom? Minority perspectives on allyship in Danish queer spaces. *British Journal of Social Psychology* 64, 1 (2025), e12780.

[112] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[113] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R Rick, and Lilly Irani. 2021. HCI tactics for politics from below: Meeting the challenges of smart cities. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–15.

[114] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* 8 (2019).

[115] Christine T Wolf. 2020. Democratizing AI? experience and accessibility in the age of artificial intelligence. *XRDS: Crossroads, The ACM Magazine for Students* 26, 4 (2020), 12–15.

[116] Nur Yildirim, Susanna Zlotnikov, Deniz Sayar, Jeremy M Kahn, Leigh A Bukowski, Sher Shah Amin, Kathryn A Riman, Billie S Davis, John S Minturn, Andrew J King, et al. 2024. Sketching ai concepts with capabilities and examples: ai innovation in the intensive care unit. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. 1–18.

[117] Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday* (2024).

[118] Cristina Zaga and Maria Luce Lupetti. 2022. *Diversity equity and inclusion in embodied AI: reflecting on and re-imagining our future with embodied AI.* 4TU. Federation.

[119] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric PS Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.