Author version of accepted full paper:
Eighth AAAI/ACM Conference on AI, Ethics, and Society (AIES ‘25)

Participatory AI and the EU AI Act

Chiara Ullstein, Simon Jarvers, Michel Hohendanner, Orestis Papakyriakopoulos, Jens Grossklags
Technical University of Munich, Germany
{chiara.ullstein, simon.jarvers, michel.hohendanner, orestis.p, jens.grossklags}@tum.de

Abstract

Participatory AI calls for the involvement of stakeholders in AI design, development, evaluation, and deployment to attain more inclusive, transparent, and accountable AI. However, actual implementations of participatory AI remain little incentivized by governments, despite appeals issued by academia and also industry. In this work, we investigate the role of ‘participation’ in the obligations of AI system providers and deployers set out by the EU AI Act. First, we analyze the gaps between the participation explicitly stated in the non-binding recitals of the AI Act and the provisions of the Act itself, showing that the legal demand for participation is limited. For example, neither Article 9 on risk management systems nor Article 27 on the fundamental rights impact assessment mention any form of participation. Article 95 on the voluntary codes of conduct is the only enacting term that explicitly suggests stakeholder participation. Second, based on these results, we analyze opportunities for participation emerging from the obligations of high-risk AI system providers and deployers (AI Act, Chapter III, Sections 2 and 3). We identify five clusters of obligations with participatory opportunities: risk management, data and data governance, information provision, resilience testing, and impact assessment. Third, we provide examples of use cases for each of the identified opportunities for participation. This work contributes to a better understanding of regulatory demands and practical opportunities regarding participatory AI in the context of the AI Act.

1 Introduction

Artificial intelligence (AI) systems are implemented across diverse domains such as hiring, healthcare, education, and content moderation, and their impact increasingly affects societies at large, with partially consequential outcomes. As a result, there has been a growing demand from various fields to involve stakeholders in the design, development, evaluation, and deployment of AI systems (Amershi et al. 2019; Birhane et al. 2022; Kulynych et al. 2020; Wolf et al. 2018). This call for participation has made way for a “participatory turn” in AI (Delgado et al. 2023), with respective activities being referred to as *Participatory AI* or *Participatory ML* (Delgado et al. 2023; Feffer et al. 2023; Birhane et al. 2022).

Fraisl et al. (2025) argue that the integration of citizen science into AI, as one specific participatory approach, could tackle significant AI-related challenges like social bias. Other scholars researching participatory AI and ML highlight that these approaches can achieve better alignment of the AI system with users’ and affected parties’ values, preferences, and needs, thereby transforming the traditional designer-user dynamic into collaborative co-design and co-creation relationships where stakeholders are empowered (Delgado et al. 2023; Feffer et al. 2023; Birhane et al. 2022; Amershi et al. 2019; Bondi et al. 2021; Denton et al. 2020).

The implementation of participatory AI, however, faces challenges from meaningless participation to high costs and issues of scale (Young et al. 2024; Groves et al. 2023; Sloane et al. 2022). Recently, regulatory efforts have tried to legally demand or motivate voluntary participatory approaches such as consultations for community input, but have remained largely ineffective (Young et al. 2024; The White House 2023; Wilson 2022).

With the aim of promoting trustworthy AI by safeguarding EU citizens against threats to their health, safety, and fundamental rights, the EU’s Artificial Intelligence Act (AI Act) by the European Parliament and The Council (2024) generally aligns with the principles of participatory AI. It represents a shift in AI governance from voluntary codes of conduct to legally binding regulation. Given the complexity and societal impact of AI systems, which necessitate governance mechanisms that go beyond traditional regulatory compliance to include meaningful stakeholder engagement, the question arises as to *whether the AI Act serves as a lever to incentivize participation in the design, development, evaluation, and deployment of AI systems.*

To study this question, (1) we analyze the participation-oriented language in the recitals of the AI Act and compare identified proposals for participation with requirements in the legal text (*gap analysis*). Finding no legally binding requirements for participation beyond informing, (2) we analyze the potential for participatory AI in the AI Act’s obligations of high-risk AI system providers and deployers (*opportunity analysis*). We identify five clusters of obligations with participatory opportunities: risk management, data and data governance, resilience testing, information provision, and impact assessment. (3) We map these to examples of participatory AI implementations.

We contribute to the field by documenting a lack of incentives in the AI Act for high-risk AI system providers and deployers to foster participation beyond informing. We show the breadth of participatory opportunities along the AI Act's high-risk AI system obligations and provide examples on how participatory AI can be applied to fulfill requirements imposed by the AI Act. This research advances the AI ethics community by providing the first systematic analysis of participatory requirements and opportunities within the AI Act, highlighting concrete participatory opportunities for practitioners navigating the intersection of participatory design principles and legal compliance requirements.

The paper proceeds as follows: Section 2 describes that participation in the design of technologies is not a new phenomenon, but that, although much conceptual work has been done on participation, there are still obstacles that have hindered its implementation despite initial regulatory efforts. Section 3 outlines the research design, followed by methods and results for the gap analysis in Section 4, and the opportunity analysis in Section 5. Section 6 presents use cases for identified participatory opportunities. Section 7 contextualizes the results concerning participation mode and incentives for participation, and addresses limitations and future research. Section 8 offers concluding remarks.

2 Background

2.1 Participation in Technology Design

In technology design, participation is most closely associated with the *Collective Resource Approach* or *Scandinavian approach*, today also known as *Participatory Design* (Asaro 2000; Basballe, Halskov, and Hansen 2016; Dearden and Rizvi 2008). This approach emerged from the labor movement of the 1970s, first, through increased activities of trade unions (*collective*) and their engagement with workers (*resource*) in the context of workplace decision-making, and later also through the development of technology according to workers' skills and interests (Asaro 2000). The approach gained prominence through the first participatory design development project UTOPIA in 1981, laying strong grounds for a tradition of cooperation between users and researchers in technology development and decision-making for the improvement of work situations (Asaro 2000; Sundblad 2010).

The methodology, rooted in social democratic values, involves joint workshops, consensus building, or building long-term relationships, among others, to reduce the power imbalance between experts and users. The approach aimed to ensure that technology serves people's (i.e., initially workers') needs while strengthening democratic legitimacy through inclusive processes (Asaro 2000; Sundblad 2010). It has significantly influenced global practices of participatory design (Basballe, Halskov, and Hansen 2016; Asaro 2000).

With the transition of Participatory Design out of Scandinavia to North America in the 1990s, participatory approaches have not only adapted but also spread more broadly into society (Basballe, Halskov, and Hansen 2016). For example, one adaptation was a shift in focus from workers to other parts of society (non-workers), such as patients or school children (Basballe, Halskov, and Hansen 2016).

2.2 Taxonomies of Participation

Beyond technology design, the concept of participation has also roots in different scholarly communities, such as international development (Dearden and Rizvi 2008; Pretty 1995), or public policy (Arnstein 1969), which have produced early but influential taxonomies of participation.

In the United States, the late 1960s and early 1970s witnessed a heated controversy around citizen participation in public policy (Arnstein 1969). The discussion centered around reasonableness and feasibility of the participation of those who are not in power (i.e., the have-nots). This is when Arnstein (1969, p.216) published her influential article *A Ladder of Citizen Participation*, and defined citizen participation as "the redistribution of power that enables the have-not citizens [...] [to] induce significant social reform which enables them to share in the benefits of the affluent society." Taking the perspective of the receivers of projects or programs (Cornwall 2008), Arnstein (1969) suggests a distinction of eight rungs across three levels of participation, including non-participation (manipulation, therapy), tokenism (information, consultation, placation), and citizen power (partnership, delegated power, citizen control). Arnstein (1969) highlights that "participation is ultimately about power and control" (Cornwall 2008, p.271).

Pretty (1995) presents a typology of participation from the perspective of users of participation, categorizing it into manipulative, passive, consultative, material incentive-based, functional, interactive, and self-mobilization; the latter largely remaining an unattainable goal (Cornwall 2008). This model emphasizes that the motivations behind participation are key to shaping interventions. White (1996) offers a typology of interests, differentiating forms of participation (nominal, instrumental, representative, transformative) by their meaning for the implementing agencies (legitimation, efficiency, sustainability, empowerment) and those on the receiving end (inclusion, costs, leverage, empowerment).

Keeping the distinction of implementing agencies (i.e., public participation goal) and receivers (i.e., promise to the public), the International Association for Public Participation (IAP2 1999) designed a widely adopted spectrum of participation with the levels inform, consult, involve, collaborate, empower. Importantly, the spectrum demonstrates that varying degrees of participation can be appropriate based on factors such as objectives, available time frames, resource constraints, and the significance of stakeholder concerns within specific decision-making contexts (IAP2 2025).

Taken together, these taxonomies provide frameworks for analyzing and conceptualizing participatory interventions, considering who is initiating participation, what the goal is, who is receiving participation, and what the promise is. In particular, the latter spectrum of the IAP2 (2025) presents the perspective that all forms of participation may be legitimate depending on the specific context.

2.3 Participatory AI Frameworks

Scholars in the AI realm have adapted, among others, the above-mentioned taxonomies and contributed to the field with conceptual frameworks for classifying differences in participatory practices in the AI context. For example, the Ada Lovelace Institute (2021) uses the same levels and the same logic of describing promises to the public as the IAP2 (2025) in their spectrum of participation in data stewardship. Berditchevskaia, Malliaraki, and Peach (2021) define the levels consultation, contribution, collaboration, and co-creation in their framework for operationalizing participatory AI for humanitarian innovation. Adopting a different viewpoint on participation (i.e., the *why*), Birhane et al. (2022) outline three goals of participatory AI: enhancing algorithmic performance, improving processes, and facilitating collective exploration. Sloane et al. (2022) differentiate between participation as work, participation as consultation, and participation as justice, and caution against potentially extractive and exploitative formats of participation.

Delgado et al. (2023) build upon this and prior work on participatory traditions from other disciplines and design a conceptual framework that maps each mode of participation (consult, include, collaborate, and own) to the dimensions participation goal (why), participation scope (what, who), and form of participation (how). Although participatory AI frameworks from within the AI community become more elaborate, there is still a lack of consensus on participatory standards, and many obstacles associated with the implementation of participatory AI remain (Birhane et al. 2022; Delgado et al. 2023; Sloane et al. 2022; Young et al. 2024).

2.4 Obstacles to Participatory AI

Challenges to implementing participation have also been reported early on. For example, Kraft and Bansler (1992) pointed out that the Collective Resource Approach had not been sufficiently accepted by workers and unions, and had little impact on practices at workplaces, prompting discussions on the need to reconsider important aspects of participatory design (Bødker 1994).

With the globalized operation and overall scale of commercial AI systems, there is a renewed need for changes in participatory approaches (Young et al. 2024). This also means understanding how the value of stakeholder input for AI developers with few resources (such as start-ups or non-governmental organizations) can be increased. Further institutional barriers include high costs and resource demands, organizational 'atomization', or absence of transparency incentives during implementation (Groves et al. 2023). Other challenges concern navigating disagreements or conflicting agendas (Kallina and Singh 2024; Ullstein et al. 2024).

There is agreement that participation should not just be an empty ritual (Arnstein 1969). Further, the criticism of participation washing is not unjustified as forms of participation can also have an exploitative effect (Sloane et al. 2022; Groves et al. 2023). This may stem not only from the form participation takes, but also from who is included or excluded – potentially reinforcing, rather than redistributing, existing power imbalances.

2.5 Participation In and Around AI Governance

The need for appropriate involvement of affected communities or their representatives in the design, development, evaluation, and deployment of AI systems has also been acknowledged in the regulatory realm (Young et al. 2024). In a review of national AI strategies from 2020, Wilson (2022, p.8) finds that references to participation and public engagement in AI governance exist in most strategies, but "were usually abstract and consistently overshadowed by other roles, values and policy concerns".

In the United States, there have been legislative efforts, such as the Algorithmic Accountability Act (introduced into Congress in June 2025, revised since 2019 (Mökander et al. 2022)), that would require developers to "meaningfully consult [...] including through participatory design" with relevant stakeholders for algorithmic impact assessments (U.S. Congress 2022). In the revoked 2023 AI Executive Order (The White House 2023, 2025), federal agencies were mandated to engage with affected communities to prevent discrimination. Other initiatives, such as the AI risk management framework by the National Institute of Standards and Technology (2023), still represent a larger push for stakeholder engagement, but do not provide legally binding incentives.

In the EU, for example, the Digital Services Act assigns "trusted flagger" organizations (awarded by the Digital Services Coordinator of the Member State) the responsibility to detect potentially illegal content and alert online platforms (Article 22, European Parliament and The Council 2022).[1] The EU General Data Protection Regulation stipulates that "[w]here appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing" as part of the data protection impact assessment (Article 35, European Parliament and The Council 2016). Referring to the context of audits, Hartmann et al. (2025) identify a regulatory gap where the AI Act does not grant data access to researchers and civil society organizations, preventing effective academic or civil society oversight through independent audits of AI systems. However, beyond this study, the need to better understand the role of participation in the AI Act remains.

Overall, this brief overview indicates that so far, legally binding AI and data governance approaches mandating companies and other impactful organizations to implement participation exist, but are rare and not comprehensive.

3 Research Design

Analysis Scope. To explore the role of participation in the design, development, evaluation, and deployment of AI systems under the AI Act, in this study, we focus on participation that occurs between providers or deployers and end users. We define end users as individuals or groups that use or work with the AI system or are subject to or affected by it. We do not consider forms of participation that take place at an (inter-)institutional level aimed at improving the design or enforcement of the AI Act itself, such as the development

[1] In the past, data cleaning work through external contractors has also proven to be implemented as an exploitative practice rather than a truly participatory mechanism (Sloane et al. 2022; Hao and Seetharaman 2023).

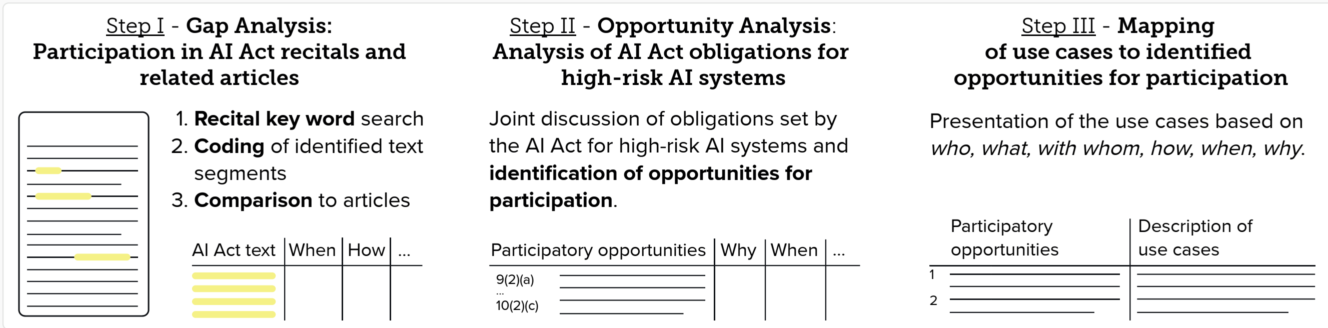

Figure 1: Research Design Procedure.

of the General-Purpose AI (GPAI) code of practice.[2]

Overview of Research Design. Our research design includes three steps, illustrated in Figure 1. First, we analyze the gaps between participation explicitly set out by the recitals of the AI Act and the enacting terms. We consider the approach suitable given that recitals in EU legislation have the function of "set[ting] out the reasons for the contents of the enacting terms (i.e. the articles) of an act" (Publications Office of the European Union 2022, p.35) and are printed before the articles. They can sometimes be considered a legally non-binding compromise if consensus is lacking (Krommendijk and Borgesius 2023). Hence, they may provide insights into extant intentions. Based on the results of this first analysis, which shows limited requirements for participation by the AI Act, we analyze opportunities for participation emerging from the requirements and obligations for high-risk AI system providers and deployers (AI Act, Chapter III, Section 2 and 3).[3] Finally, we provide examples of use cases for each of the identified opportunities for participation.

Methodologically Related Prior Work. Wilson (2022) studies the public's role and public engagement mechanisms in 16 national AI strategies through content analysis of these national strategy documents. For this analysis, he identifies relevant text portions through text search queries (drawn from literature discussed in the background section of his work). The text portions were then hand-coded in three rounds, focusing on the different research questions.

Rebrean and Malgieri (2025) study the concept of vulnerability in the EU AI Act, offering a comprehensive interpretation of vulnerability references in the AI Act. They investigate explicit references of, e.g., "vulnerabilities" and "vulnerable" in the recitals and articles to formulate a definition of the concept of vulnerability as used in the AI Act. Rebrean and Malgieri (2025) emphasize that understanding the examples that the AI Act associates with each of the references is an important step in formulating their definition.

We are not aware of other research taking a comparable analysis approach. In the following, we outline our research design for the gap analysis (Step I) and opportunity analysis (Step II) and present our findings for each of the analyses.

4 Step I: Gap Analysis of Participation in AI Act Recitals and Related Articles

4.1 Methods: Gap Analysis

Data Collection: AI Act Recital Keyword Search. To identify text portions of the AI Act relevant to participation, we performed a keyword search within the recitals of the AI Act. We selected keywords pertaining to participation and stakeholders drawn from the literature discussed in the Background Section and iteratively refined the selection: *stakeholder, civil society, affected, vulnerable, public interest, expert, human, discriminat, inclusive, participat, involvement, input, engagement, consult, inform*. The keyword search resulted in the identification of 101 recitals, whereby one recital could contain multiple keywords. We defined the following inclusion and exclusion criteria for consideration of a recital for further analysis (see Table 1 for examples): We included a recital if it (1) referred to *participation between provider/deployer and end user* (vs. participation between provider and deployer), and described (2) *actionable participation mechanisms*, or (3) *specified diverse stakeholders*. We excluded a recital if it described (1) *abstract principles without mechanism*, (2) the *applicability of regulation*, (3) *inter-institutional arrangements or procedural mechanisms*, or (4) *technical specifications of the system's functioning*, or used (5) *passive protection language or described harm*. Multiple criteria could be assigned to a recital. We only analyzed those sentences of a recital that fell into the inclusion criteria. Twelve recitals met the inclusion criteria, referring to participation in design, develop-

[2]These procedural participatory mechanisms in the AI Act are out of the scope of this paper, as, at the time of our analysis, some of these mechanisms were ongoing (e.g., GPAI code of practice) or had not yet commenced (e.g., public consultations on high-risk AI systems), limiting the analysis of their implementation.

[3]We focus on all obligations for high-risk AI systems, as they have only been partially covered in the identified recitals/articles in the gap analysis, and participation may have a great impact on their development. We exclude the obligations of GPAI providers (Articles 53-55) from the opportunity analysis because the GPAI code of practice (Article 56), a voluntary tool that helps comply with the obligations, had not yet been published at the time of our analysis. Nevertheless, the results from the opportunity analysis may remain relevant, as the obligations of providers of GPAI (with systemic risk) also align with the identified five clusters of participatory AI opportunities.

Criteria		Recital	Quote
Inclusion	(1)	Recital 96	deployers of high-risk AI system [...] could involve relevant stakeholders
	(2)	Recital 96	involve [...] in conducting such impact assessments and designing measures
	(3)	Recital 96; Recital 171	independent experts, and civil society organisations; affected persons
Exclusion	(1)	Recital 1	human centric [...] AI
	(2)	Recital 21	rules [...] should apply [...] in a non-discriminatory manner
	(3)	Recital 149	The Board should also cooperate [...] with [...] experts
	(4)	Recital 12	AI systems [...] have some degree of independence of actions from human involvement
	(5)	Recital 32	entail discriminatory effects

Table 1: Examples for Inclusion and Exclusion Criteria

ment, evaluation or deployment of AI systems: Recitals 20, 27, 65, 92, 93, 96, 131, 132, 134, 141, 165, 171.[4] To address the limitation that keyword searches may miss relevant language, we manually reviewed all remaining 79 recitals and found no additional recitals meeting our inclusion criteria.

Data Analysis. First, a researcher applied the inclusion and exclusion criteria. This analysis was validated by a second researcher, where two differences in coding were discussed; both recitals were excluded. Then, two researchers independently coded the data set regarding the categories *who* should initiate participation about *what*, with *whom*, *how*, and *when* in the AI lifecycle, and *why* participation could be beneficial. The categories why, what, with whom, and how were derived from the framework by Delgado et al. (2023). By also including the *who* as implementing agency, our analysis captures the distinction between implementing agencies that pursue certain goals and the receivers of participation (White 1996; IAP2 2025).

Based on the taxonomies and frameworks for participation presented in the Background Section, for the category how, we distinguished between "inform" and "consult and involve" (IAP2 2025). Although we originally intended to further differentiate the forms of participation mentioned in the AI Act, we soon realized that there was a lack of detailed specification of the forms of participation, with references to collaboration or empowerment being completely absent.

We coded all categories (who, what, with whom, how, when, why) inductively to avoid losing information from the recitals. After individual coding (first coding wave), we discussed the identified codes for each recital and category, and re-coded the data based on the discussed codes (second wave of coding). The result of this coding procedure was jointly discussed again. We then jointly compared the results with the legal text of the related articles[5] of the AI Act.

[4]Note that some sentences of Recitals 20 and 165 also relate to the exclusion criteria (3) inter-institutional arrangements/ procedural mechanisms.

[5]We refer to the related articles as referenced by the Future of Life Institute (2025) for each of the recitals, if mentioned.

4.2 Results: Gap Analysis

Topics of Identified Recitals. We identified twelve recitals that addressed some form of participation between the provider or deployer of AI systems and the end user (or their representatives). These recitals consider AI literacy promotion (Recital 20), ethical principles for AI (Recital 27), risk management system requirements (Recital 65), information of workers and persons (Recitals 92 and 93), fundamental rights impact assessments (Recital 96), EU database (Recital 131), AI system transparency requirements and deepfakes (Recital 132 for provider perspective and Recital 134 for deployer perspective), testing in real world conditions (Recital 141), voluntary codes of conduct (Recital 165), and the right to obtain an explanation (Recital 171). Figure 2 maps these recitals based on *who*, with *whom*, and *how*.

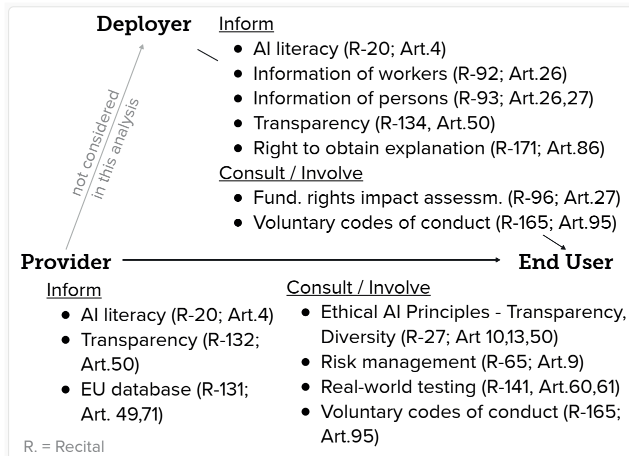

Figure 2: Classification of Identified Recitals by *Who* (Provider/ Deployer), with *Whom* (End User), and *How* (Inform or Consult/ Involve).

Mode of Participation Suggested in Recitals. Eight of the twelve identified recitals suggest *informing* as participation mode (Recitals 20, 27, 92, 93, 132, 131, 134, 171). They highlight that "affected persons" (Recitals 20, 27, 171), "humans" (Recital 27), "relevant" (Recital 20) or "diverse actors" (Recital 27), "workers" (Recital 92), "natural persons" (Recitals 93, 132), or "vulnerable groups" (Recital 132) should be made "aware" (Recital 27), "be notified" (Recital 132) or "inform[ed]" (Recital 93), "obtain an explanation" (Recital 171), "find relevant information" (Recital 131), or be "equipped [...] with the necessary notions to make informed decisions" (Recital 20). The time point is often not specified; mostly, it refers to the deployment phase.

Five of the twelve identified recitals suggest participation that goes beyond informing, i.e., *consulting and involving* (Recitals 27 (also *informing*), 65, 96, 141, 165). However, these recitals remain vague on the form of participation. Recital 27, referring to ethical principles for AI, suggests "that AI systems are developed and used in a way that includes diverse actors" contributing to diversity, non-discrimination, and fairness. Recital 65, referring to risk management system requirements, suggests that "[w]hen identifying the most appropriate risk-management measures, the provider should [...], when relevant, involve experts and external stakeholders." Recital 96, referring to fundamental rights impact assessments, suggests that "to collect relevant information necessary to perform the impact assessment, deployers of high-risk AI system [...] could involve relevant stakeholders [...]." Recital 141, referring to testing in real world conditions, suggests that "natural persons [...] participate in testing in real world conditions [...]." Most explicitly, Recital 165, referring to codes of conduct, suggests "stakeholders' participation with the involvement, as appropriate, of relevant stakeholders [...]."

Comparison of Participation Suggested in Recitals and Enacting Terms. Most enacting terms that refer to any of the abovementioned recitals reflect the participation mode of *informing*. In particular, the transparency obligations of Article 50, which require providers and deployers to inform the concerned natural persons "that they are interacting with an AI system", reflect the *inform* mechanisms referenced to in Recitals 27, 93, 132, and 134. Similarly, the information requirements of Article 26 (Recital 92) necessitate deployers to inform workers "that they will be subject to the use of the high-risk AI system." In contrast, Article 86 on the right to explanation of individual decision-making reflects Recital 171 but does not explicitly mention that "deployer[s] should also inform the natural persons about their right to an explanation" (Recital 93). Neither does Article 4 on AI literacy explicitly mention that "relevant actors in the AI value chain" (Recital 20) should be informed or equipped with knowledge. While Recital 131 specifies one purpose of the EU database as "allowing the general public to find relevant information," Articles 49 and 71 focus on the details of the EU database for high-risk AI systems as a framework.

Referring to the participation mode of *consulting and involving*, neither Article 9 on risk management systems (Recital 65) nor Article 27 on the fundamental rights impact assessment (Recital 96) mention any form of participation. Article 9 highlights the necessity to consider whether the "high-risk AI system is likely to have an adverse impact on persons under the age of 18 and, as appropriate, other vulnerable groups" and Article 27 requires an assessment of "the specific risks of harm likely to have an impact on [...] natural persons or groups of persons." Neither does Article 60 on testing of high-risk AI systems in real world conditions outside AI regulatory sandboxes specify participation, but it addresses safeguards for data protection and informed consent, e.g., "the subjects of the testing in real world conditions who are persons belonging to vulnerable groups due to their age or disability, are appropriately protected" (Article 60(4)(g)). The suggestion of including diverse actors in the development of AI systems (Recital 27) may refer to Article 10 on data and data governance, which requires data to be "sufficiently representative" (Article 10(3)) and take into account the "characteristics or elements that are particular to the specific geographical [...] setting within which the high-risk AI system is intended to be used" (Article 10(4)). However, again, participation is not explicitly addressed.

The only article that explicitly puts forward stakeholder participation as set out by the related recital is Article 95 on the codes of conduct (Recital 165). The article addresses participation as a means of AI design: "facilitating an inclusive and diverse design of AI systems, including through the establishment of inclusive and diverse development teams and the promotion of stakeholders' participation in that process" (Article 95(2)(d); as well as for the process of drawing up the codes of conduct (Article 95(3)).

Summary of Gap Analysis. We identified twelve recitals that propose participation between providers or deployers and end users through *informing* or *consulting and involving*. Most recitals (Recitals 27, 93, 132, 134), suggesting participation through informing, referred to Article 50 on the transparency obligations, which explicitly requests this level of participation. Instead, recitals suggesting participation through consulting and involving (Recitals 27, 65, 96, 141, 165) were not accordingly reflected in the associated enacting terms. For example, neither Article 9 on risk management systems nor Article 27 on the fundamental rights impact assessment mention any form of participation. Based on our analysis resulting from Step I of our research design, Article 95 on the voluntary codes of conduct is the only enacting term that explicitly suggests stakeholder participation. However, this article describes voluntary activities and is not part of the obligations for high-risk AI system providers and deployers. For this reason, in Step II of our research design, we review all *obligations of providers and deployers of high-risk AI systems* for implicit indications of and opportunities for participation.

5 Step II: Participatory AI Opportunity Analysis in High-Risk AI System Obligations

5.1 Methods: Participatory AI Opportunity Analysis

Data Collection: AI Act Ch. III, Sec. 2 & 3 Legal Text. AI Act Chapter III, Section 3 describes all *obligations of providers and deployers of high-risk AI systems and other parties*. It specifies the *obligations of providers of high-risk AI systems* in Article 16 and the *obligations of deployers of high-risk AI systems* in Article 26. Article 16(a) requests compliance "with the requirements [for high-risk AI systems] set out in Section 2." Hence, for Step II of the research design, two researchers individually reviewed AI Act Chapter III, Section 2 and Section 3 for implicit indications of and opportunities for participation. We defined an article to contain an *opportunity for participation* (i.e., Why?) if it can be argued that stakeholder involvement can contribute knowledge or artifacts that the AI provider or AI deployer team lacks. The same two researchers compared their individually identified legal text elements and discussed three disagreements, of which two were dropped and one was kept. This led to a data set of 29 snippets containing legal text from Articles 4, 9, 10, 13 - 17, 26, 27, and 72.

Analysis. The two researchers discussed each of the 29 snippets, grouped them into clusters of opportunities, and formulated the *participatory opportunities*. Then, they jointly coded the snippets to identify *who* initiates participation *with whom* (participants) on *what* and *when* (approximate AI lifecycle step). The results of this analysis are presented in Table 2, in which the clusters are ordered according to their prototypical occurrence along the AI lifecycle. We acknowledge the limitation that multiple cluster topics do not only occur at one point in time of the AI lifecycle, but are recurring or of relevance to multiple AI lifecycle steps. We mark these instances with footnotes.

5.2 Results: Participatory AI Opportunity Analysis

Summary of Obligations Granting Opportunities for Participation. We identified five clusters of high-risk AI system obligations that contain opportunities for participation. The clusters are illustrated in Table 2 with five different colors: risk management (Article 9), referencing accessibility requirements (Article 16(l)) and post-market monitoring (Article 72); data and data governance (Article 10); informatory obligations including human oversight (Article 14), transparency and instructions for use (Article 13), quality management system (Article 17), AI literacy (Article 4) and certain deployer obligations (Article 26); resilience testing (Article 15); and impact assessment (Article 26 and 27). The majority of obligations refer to AI system providers, except for the obligations to perform a fundamental rights impact assessment, to inform when subject to high-risk AI systems, to collect relevant data as part of the post-market monitoring system, and to provide AI literacy, which refer to high-risk AI system deployers.

Risk Management. For each part of the risk management system, participation can be considered relevant. While participation at the stage of problem formulation refers more to understanding, defining, or speculating about concepts, such as what constitutes "foreseeable risks" or forms of "misuse", participation at later stages, such as testing and evaluation, allows for performing risk assessments with affected individuals, their representatives, or domain experts. The risk management system also plays an important role during deployment and monitoring, where participation can contribute to collecting data and monitoring or evaluating newly arising risks.

Data and Data Governance. The participation of stakeholders constitutes an important foundation for effective data management frameworks, in particular, concerning data collection and examination of data. Stakeholders, such as affected individuals or people with subject matter expertise, can help determine whether the data are, e.g., representative regarding the intended purpose, the AI subjects, or geography, context, behavior, and function. In the context of preprocessing the data, stakeholders can help identify data gaps and biases and enhance the definition of concepts for preprocessing or the procedures for renewed data collection.

Resilience Testing. The cluster of resilience testing grants, in particular, the potential for red teaming. Participation of stakeholders, e.g., in the form of red teaming, can facilitate assessing robustness or cybersecurity. Furthermore, appropriate levels of robustness and cybersecurity can be determined collaboratively.

Information Provision. The participation of relevant stakeholders can improve the information provided as part of the documentation and transparency requirements. Deployers, customers, or workers can provide feedback on, e.g., whether the information provided is sufficient and clear or whether the human-machine interface is an effective tool for overseeing the AI system. The latter opportunity can be related to user testing. Furthermore, deployers are required to inform people if they are subject to high-risk AI systems.

Impact Assessment. Finally, the cluster of impact assessments grants the opportunity to collaboratively perform assessments of the impact on fundamental rights or data protection, if applicable. Also, governance measures and complaint mechanisms for risk scenarios can be developed jointly with affected individuals, their representatives, or domain experts.

6 Step III: Mapping of Use Cases to Identified Participatory Opportunities

Through a *gap analysis* (Step I), this paper has shown that, according to the AI Act, there is no legal obligation for participation beyond that of 'informing' relevant people, although some recitals advocate 'consulting or involving' stakeholders. Based on this analysis, through an *opportunity analysis* (Step II), we identified five clusters of high-risk AI obligations granting opportunities for participation along the AI lifecycle. In the following, we map examples of participatory AI use cases to each of the five identified opportunities for participation (Step III). Similarly to the approach by Corbett, Denton, and Erete (2023) to presenting use cases, our aim was to identify examples of participatory AI realizations to illustrate how participation could take place for different AI Act requirements. We describe each use case based on the analysis criteria *who, what, with whom, how, when, why* and briefly reflect on each effort.

Data and Data Governance. To demonstrate the collection of relevant and representative data that takes into account geographical and contextual nuances (Art. 10(3/4)), we refer to the machine translation project by the researchers Nekoto et al. (2020) (*who*). Motivated by the low-resource language problem in machine translation, which is not solvable by researchers alone, they involved all the necessary agents required in the machine translation development process (*why*). The project produced new translation datasets and machine translation benchmarks for more than 30 African languages, with human assessments performed for approximately one third of these languages (*what*). More than 400 participants from at least 20 countries without formal training (contacted through multiple communication channels) contributed to the project (*with whom*). For example, to collect data (*when*), participants self-organized to translate writings in collaborative sessions (*how*).

[6] also: Problem Formulation / Testing and Evaluation
[7] also: Testing and Evaluation
[8] also: Testing and Evaluation
[9] also: Problem Formulation
[10] also: Problem Formulation
[11] also: Training and Validating
[12] Most of the referenced articles in this lifecycle step relate more to testing and evaluation than to design and training.

Lifecycle		Referenced Topic	Article	What?	Why?	Who?	With Whom?
General	Risk Management	Risk Management[6]	9(9)	Consideration of adverse impacts	Understanding potential adverse impacts on vulnerable groups	P	Persons < 18, vulnerable groups, representatives, experts
		↪ Obligations of Providers[7]	16(l)	Compliance with other accessibility requirem.	Determining appropriate accessibility measures	P	Affected, representatives, domain experts
Problem Formulation		Risk Management	9(2)(a)	Risk identification & analysis (intended use)	Determining what "foreseeable" risks means, and how to identify and analyze	P	Affected, representatives, domain experts
		Risk Management[8]	9(2)(b)	Risk estimation & evaluation ((un)intended use)	Determining what "misuse" means, how to estimate/evaluate	P	Affected, representatives, domain experts
Data Collection	Data and Data Governance	Data & Data Governance	10(3)	Relevant, representative, complete data	Ensuring data representativeness and collection of such data	P	Target user groups, data providers
		Data & Data Governance	10(4)	Geographical, contextual, behavioral, functional representativeness	Understanding specific nuances of settings	P	Affected, representatives, domain experts, data providers
Data Preprocessing		Data & Data Governance	10(2)(h)	Identification of relevant data gaps & limitations	Identifying data gaps & limitations, and mitigation strategies	P	Affected, representatives, domain experts
		Data & Data Governance	10(2)(f)	Examination of biases	Examining potential biases	P	Affected, representatives, domain experts
		↪ Data & Data Governance	10(2)(g)	Bias detection, prevention & mitigation	Identifying measures to detect, prevent, and mitigate biases	P	Affected, representatives, domain experts
		Data & Data Governance	10(2)(c)	Data-preparation processing operations	Defining pre-processing concepts (e.g., labeling, updating)	P	Affected, representatives, domain experts
Design, Training, Testing and Evaluation[12]	Risk Management	Risk Management[9]	9(2)(d)	Risk management measures	Determining appropriate risk management measures	P	Affected, representatives, domain experts
		↪ Risk Management[10]	9(5)	Evaluation of acceptable residual risk	Determining what residual risk can be judged acceptable	P	Affected, representatives, domain experts
		↪ Risk Management	9(6)	Identification of risk management measures	Testing of AI systems to identify best risk mgmt. measures	P	Affected, representatives, domain experts
		↪ Accuracy, Robustness, Cybersecurity[11]	15(1)	Identification of appropriate accuracy levels	Testing AI and determining "appropriate" accuracy level	P	Affected, representatives, domain experts
	Resilience Test.	Robustness	15(4)	Resilience testing (system robustness)	Testing of AI system, e.g., through red teaming	P	Domain experts
		Cybersecurity	15(5)	Resilience testing (cybersecurity)	Testing of AI system, e.g., through red teaming	P	Domain experts
		Accuracy, Robustness and Cybersecurity	15(1)	Appropriate robustness & cybersecurity levels	Determining "appropriate" levels	P	Affected, representatives, domain experts
	Information Provision	Human Oversight	14(1)	Design of interfaces for oversight	Obtaining feedback on effectiveness of oversight tool	P	Operative deployers overseeing AI system
Pre-preparation for Deployment		Transparency	13(1)	Sufficiently transparent operation	Obtaining feedback on transparency for interpretability	P	Operative deployers using AI system
		Instructions for Use	13(2)	Provision of instructions	Obtaining feedback on comprehensibility of instructions	P	Operative deployers using AI system
		QMS	17(1)(j)	Communication channels	Establishing stakeholder communication	P	Customers, other interested parties
		AI Literacy	4	Measures for AI literacy	Understanding needs of AI system operators and users	P, D	Persons operating or using the system
		Obligations of Deployers	26(7)	Information about being subject to high-risk AI	Informing workers and representatives about high-risk AI	D	Workers and representatives
		Obligations of Deployers	26(11)	Information about being subject to high-risk AI	Informing natural persons about subject to high-risk AI	D	Persons subject to high-risk AI
	Impact Assess.	Fundamental Rights Impact Assessment	27(1) / 27(1)(c,d)	Assessment of impact on fundamental rights	Assessing affected groups and impacts	D	Affected, representatives, domain experts
		Fundamental Rights Impact Assessment	27(1)(f)	Identification of mitigation measures	Developing governance and complaint mechanisms for risk	D	Affected, representatives, domain experts
		Obligations of Deployers	26(9)	Data protection impact assessment, if applicable	Assessing data protection operations	D	Affected, representatives, domain experts
Deploy. & Monitor.	Risk M.	Risk Management	9(2)(c)	Risk evaluation during deployment	Determining what "other risks" means, how to estimate/evaluate	P	Affected, representatives, domain experts
		↪ Post-market Monitoring	72(2)	Collection, document., and analysis of data	Conducting continuous AI performance monitoring	D	Affected, representatives, domain experts

Table 2: AI Act obligations along the AI lifecycle granting opportunities for participation. P: provider; D: deployer.

Birhane et al. (2022) emphasize the project's grassroots participatory nature with genuine empowerment through reciprocal, bi-directional processes where participants shape core decisions around the data and the benchmarks. In a context of open sourcing the created data, Birhane et al. (2022) highlight the risk of co-optation by commercial actors who may exploit the participatory data and tools for profit without supporting the broader community effort.

Resilience Testing. Participation in AI system testing and evaluation (Art. 15(4); *when*) can be realized through structured public challenges, as demonstrated by Anthropic's Constitutional Classifiers initiative ((*use case*); Sharma et al. 2025). This initiative engaged 405 participants, including "academic researchers, university professors, experienced LLM red-teaming contractors, and motivated newcomer[s]" (Sharma et al. 2025, p.7) recruited through HackerOne (*with whom*). The task was to break the model under experimental conditions to test its robustness, more specifically, to identify universal jailbreaks in AI systems, where outputs achieving at least half the score of helpful-only responses were considered compromised (*what*). The human red teaming exercise operated as a bug-bounty system, compensating participants up to $15,000 for a successful universal jailbreak (*how*). This initiative showcases how Antrophic, as an AI system provider and participation implementing actor (*who*), systematically involved external stakeholders in testing and evaluation phases, leveraging diverse expertise and perspectives to identify vulnerabilities.

Considering red teaming as a participatory measure must also be viewed with a critical eye. Feffer et al. (2024) identified the purposes of red teaming practices to be often vague. In the above-described example, the reason for the red teaming activity (*why*), e.g., in comparison to an internal evaluation, is not explicitly described (Sharma et al. 2025). From the participant's perspective, compensation for participation in commercial red teaming activities might be the main motivation to participate. Thereby, compensation must align with the requested resources, such as time investment (Akgul et al. 2023). The participatory notion of red teaming becomes questionable if the intervention resembles mere outsourcing of technical labor, which may also be the case in the above-described example, where participants were paid up to $15,000 for finding vulnerabilities. Another factor to consider is the emerging trend of red teaming automation, prioritizing efficiency and cost savings; a trend that has also been observed in content moderation. Researchers are cautioning against this practice given the importance of human expertise in red teaming (Zhang et al. 2025).

Information Provision. The researchers Lee et al. (2019) in collaboration with the non-profit 412 Food Rescue (*who*) developed an algorithmic donation allocation system. One component was the development of an explanation and decision support interface (*what*) to support human decision-makers who match recipients with donations. We use this case as an example for developing a human-machine interface for human oversight (Art. 14(1)), even if it does not meet all specified requirements. The researchers conducted one-hour studies that included walkthroughs and semi-structured interviews (*how*) to evaluate (*when*) how the information and explanations provided by the interface were perceived by different stakeholders, including representatives from the non-profit organization, from the recipient organizations, from the donor organizations, and volunteers (*with whom*). Stakeholder participation was deemed essential to collectively address the equity-efficiency trade-off in matching algorithms, as involving all parties in finding solutions increases their motivation to continue using the service by ensuring that their needs are respected (*why*).

Feffer et al. (2023) positively evaluate this use case, ranking high on most of their ten axes for assessing participatory AI. They highlight that participants were meaningfully and iteratively engaged throughout the algorithm development project via paid, face-to-face sessions with appropriate context and control over development processes, though this approach was resource-intensive (Feffer et al. 2023).

Risk Management and Impact Assessment. Given that risk assessment is often considered a step in impact assessments (Mantelero 2024; Bogucka et al. 2024), we provide qualitative and quantitative participatory examples that could be applied to both opportunity clusters, e.g., to understand adverse impacts of the AI system in general (Art. 9(9)), or also in preparation for deployment or for monitoring purposes (Art. 72(2); *when*). For example, the researchers Hohendanner et al. (2025) and Ullstein, Hohendanner, and Grossklags (2025) (*who*) initiated a dialogue series with citizen across Nigeria, Japan, India, Bolivia, Mexico and Germany (*with whom*) to explore the (non-)application of genAI and facial processing technologies in a desirable future from the local perspective of citizen (*why*). Part of the dialogues (*how*) was the mapping of potential consequences, the selection of the positive and negative consequences most worthy of discussion, and the evaluation of their impacts (*what*).

Researchers from The Collective Intelligence Project (2023) in collaboration with OpenAI (*who*) provide an alternative, rather quantitative approach to identifying and prioritizing risks from large language models (*what*). They involved 1000 demographically representative US-Americans (*with whom*) in a survey study (*how*), with six participants attending a follow-up round table with OpenAI. The aim was to explore public values and viewpoints on the most significant risks and harms associated with AI (*why*).

Concerning the participants, these two examples also show the challenges of achieving adequate representation of populations when studying risks associated with globally deployed AI technologies, as also noted by the authors. One key to participatory approaches, in general, is adequate representation of relevant stakeholder groups, which has been observed to be lacking in participatory AI projects (Feffer et al. 2023). Hence, participatory efforts initiated by companies should always start with the identification of all stakeholders and define and communicate the purpose of participation to address the needs of stakeholders (Pretty 1995).

7 Discussion

7.1 Contextualization of Findings

About the Value of *Informing* as Participation Mode. Our gap analysis showed a strong focus on the participation mode *inform*. In the past, some scholars have perceived higher levels of participation to be more valuable. In Arnstein's (1969) taxonomy, the provision of information to the public (third lowest rung) maps to the participation level tokenism. Cornwall (2008, p.272) highlights the normativity of taxonomies of participation and points out that depending on the context lower levels of participation can also have positive impacts: "Participation through information sharing, for example, might limit more active engagement. But it could be argued that transparency over certain kinds of information opens up the possibility of collective action in monitoring the consistency of rhetoric with practice."

In fact, a cornerstone of compliance with the AI Act is to create a technical documentation, in which required practices are documented in writing, "containing information which is necessary to assess the compliance of the AI system with the relevant requirements and facilitate post market monitoring" (Recital 71). Our opportunity analysis also identified the provision of information and monitoring as two of the five clusters that grant opportunities for participation. The documentation and informational requirements are aimed at providing institutions that have access to the information with a better information basis for the evaluation and appropriate use of the AI systems. However, as analyzed by Hartmann et al. (2025), academia and civil society have no access to the information and, therefore, no external and independent audit can be performed. This limits the possibilities of collective action (Cornwall 2008) in monitoring the AI system to observable information or to the details about which companies choose to inform the public. Prior research on audits has highlighted a lack of transparency as a major barrier to external audits, revealing imbalances in power (Radiya-Dixit and Neff 2023).

Missed Opportunity for Incentivizing Stakeholder Engagement. Our findings are in line with prior research on public engagement in national AI strategies, where "engagement rhetoric is common, references to specific engagement mechanisms and activities are rare" (Wilson 2022, p.1). Also in the recitals of the AI Act, there are references to participation in AI design, development, evaluation, and deployment through informing and involving; however, there are no specific requirements in the legally binding enacting terms of the AI Act that go beyond informing. Thus, the AI Act represents a missed opportunity to provide a legal infrastructure that demands social infrastructures (Young et al. 2024) for stakeholder engagement. As technology companies have not yet managed to successfully establish lasting mechanisms for local stakeholder input and governance across international markets, frequently leading to negative consequences (Young et al. 2024), the AI Act could have provided the incentives for investing in such infrastructure.

7.2 Limitations and Future Research

This research was focused on the participation that occurs between providers or deployers and end users in the context of AI system design, development, evaluation, and deployment. The primary aim of taking this perspective was to link the participatory AI discourse to the requirements set out by the EU AI Act. However, it leaves out areas for participation that are not central to the legal text, such as problem formulation (Martin Jr. et al. 2020), or procedurally mandated participatory mechanisms.

While multiple researchers were involved in the analysis, the presented analysis is subject to the researchers' reading of the AI Act. Further, while the keyword search for the gap analysis is based on participation literature and has been iteratively refined and manually validated, some participatory intentions in the recitals might have been overlooked. Still, identifying additional participatory intentions in the recitals would not change the finding that a gap exists between the participation-oriented language in the recitals and the legally binding requirements.

Future research could extend the opportunity analysis to the voluntary codes of conduct (Article 95), the transparency obligations (Article 50), and the obligations of GPAI providers in combination with the GPAI code of practice (Articles 53-56). Future work could also map participatory (inter-)institutional arrangements and procedural mechanisms related to the development and enforcement of the AI Act to analyze the nature and effectiveness of their implementation, e.g., the multi-stakeholder process for developing the GPAI code of practice (Article 56; European Commission 2024) or the voluntary codes of conduct (Article 95).

The AI Act is, at the time of writing this article, the only regulation specifically targeting AI systems placed on or put into service in the European market. However, there are other frameworks influencing AI governance within companies, such as standards, for example, the ISO/IEC 42001:2023 (Information Technology – Artificial Intelligence – Management System) (ISO/IEC 2023). Future research could analyze the conceptualization of participation in these frameworks that specifically intend to set standards for certain practices.

8 Concluding Remarks

In this paper, we showed that some participation-oriented language in the recitals of the AI Act exists; however, there are no legally binding requirements for participation beyond informing specified (groups of) people and beyond suggestions for stakeholder participation in the voluntary codes of conduct. While *informing* constitutes the most frequently referenced mode of participation, the information provided through these requirements is generally not available to the public, granting little benefit for external collective monitoring action. We argue that the AI Act represents a missed opportunity to incentivize stakeholder participation in AI design, development, evaluation, and deployment. In view of companies' efforts to align their procedures for compliance with the AI Act, we identify five clusters of obligations with participatory potential (risk management, data and data governance, resilience testing, information provision, and impact assessment) along the AI Act's high-risk AI system requirements. We further provide examples for realizations of participatory AI for each of the clusters. We recommend that regulatory stakeholders include more structural incentives for participation in frameworks currently under development, such as the European harmonized standards, which will support compliance with the AI Act.

Positionality Statement

Our research team brings diverse perspectives to this study through its multi-gender and cross-disciplinary composition, spanning computer sciences, political sciences, design, and privacy economics. This collaborative group, comprising both graduate researchers and faculty members, enabled us to analyze and interpret data through complementary disciplinary lenses particularly relevant to AI regulatory frameworks. The team has expertise in conducting participatory research on technology design and governance (Bridges, Appel, and Grossklags 2012; Engelmann, Herzog, and Grossklags 2020; Hohendanner et al. 2024a,b, 2025; Ullstein et al. 2024, 2025), in analyzing AI systems and their data (Andrews et al. 2023; Hirota et al. 2024; Papakyriakopoulos and Mboya 2023; Zhao et al. 2024), and in teaching on AI and the EU AI Act.

Acknowledgments

We thank the anonymous reviewers for their constructive comments. Simon Jarver's research position was financed by a Friedrich Schiedel Fellowship awarded to Chiara Ullstein. The authors acknowledge the financial support provided by the fellowship, which was instrumental in enabling this research. The fellowship is hosted by the TUM School of Social Sciences and Technology and the TUM Think Tank at the Munich School of Politics and Public Policy.

References

Ada Lovelace Institute. 2021. Participatory data stewardship. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf.

Akgul, O.; Eghtesad, T.; Elazari, A.; Gnawali, O.; Grossklags, J.; Mazurek, M. L.; Votipka, D.; and Laszka, A. 2023. Bug hunters' perspectives on the challenges and benefits of the bug bounty ecosystem. In *32nd USENIX Security Symposium*, USENIX Security '23, 2275–2291. Anaheim, CA: USENIX Association.

Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P. N.; Inkpen, K.; Teevan, J.; Kikin-Gil, R.; and Horvitz, E. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–13. New York, NY, USA: ACM.

Andrews, J.; Zhao, D.; Thong, W.; Modas, A.; Papakyriakopoulos, O.; and Xiang, A. 2023. Ethical considerations for responsible data curation. *Advances in Neural Information Processing Systems*, 36: 55320–55360.

Arnstein, S. R. 1969. A ladder of citizen participation. *Journal of the American Institute of Planners*, 35(4): 216–224.

Asaro, P. M. 2000. Transforming society by transforming technology: The science and politics of participatory design. *Accounting, Management and Information Technologies*, 10(4): 257–290.

Basballe, D. A.; Halskov, K.; and Hansen, N. B. 2016. The early shaping of participatory design at PDC. In *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops - Volume 2*, PDC '16, 21–24. New York, NY, USA: ACM.

Berditchevskaia, A.; Malliaraki, E.; and Peach, K. 2021. Participatory AI for humanitarian innovation. https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf.

Birhane, A.; Isaac, W.; Prabhakaran, V.; Diaz, M.; Elish, M. C.; Gabriel, I.; and Mohamed, S. 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*, EAAMO '22. New York, NY, USA: Association for Computing Machinery.

Bogucka, E.; Constantinides, M.; Šćepanović, S.; and Quercia, D. 2024. Co-designing an AI impact assessment report template with AI practitioners and AI compliance experts. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '24, 168–180. Washington, DC, USA: The AAAI Press.

Bondi, E.; Xu, L.; Acosta-Navas, D.; and Killian, J. A. 2021. Envisioning communities: A participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 425–436. New York, NY, USA: Association for Computing Machinery.

Bridges, F.; Appel, L.; and Grossklags, J. 2012. Young adults' online participation behaviors: An exploratory study of Web 2.0 use for political engagement. *Information Polity*, 17(2): 163–176.

Bødker, S. 1994. Creating conditions for participation: Conflicts and resources in systems design. In *Proceedings of the Participatory Design Conference*, PDC '94, 13–20. Palo Alto, CA, USA: Computer Professionals for Social Resp.

Corbett, E.; Denton, E.; and Erete, S. 2023. Power and public participation in AI. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23. New York, NY, USA: Association for Computing Machinery.

Cornwall, A. 2008. Unpacking 'participation': Models, meanings and practices. *Community Development Journal*, 43(3): 269–283.

Dearden, A.; and Rizvi, H. 2008. Participatory IT design and participatory development: A comparative review. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, PDC '08, 81–91. Bloomington, IN, USA: Indiana University.

Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 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*, EAAMO '23. New York, NY, USA: Association for Computing Machinery.

Denton, R.; Hanna, A.; Amironesei, R.; Smart, A.; Nicole, H.; and Scheuerman, M. K. 2020. Bringing the people back in: Contesting benchmark machine learning datasets. In *Participatory Approaches to Machine Learning, ICML 2020 Workshop*.

Engelmann, S.; Herzog, L.; and Grossklags, J. 2020. Should users participate in governing social media? Philosophical and technical considerations of democratic social media. *First Monday*, 25(12).

European Commission. 2024. AI Act: Participate in the drawing-up of the first General-Purpose AI Code of Practice. https://web.archive.org/web/20250728013954/https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice.

European Parliament and The Council. 2016. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679. Accessed: 2025-05-07.

European Parliament and The Council. 2022. REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065. Accessed: 2025-05-07.

European Parliament and The Council. 2024. REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. Accessed: 2024-12-27.

Feffer, M.; Sinha, A.; Deng, W.; Lipton, Z.; and Heidari, H. 2024. Red-teaming for generative AI: Silver bullet or security theater? In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '24, 421–437. Washington, DC, USA: The AAAI Press.

Feffer, M.; Skirpan, M.; Lipton, Z.; and Heidari, H. 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*, AIES '23, 38–48. New York, NY, USA: ACM.

Fraisl, D.; See, L.; Fritz, S.; Haklay, M.; and McCallum, I. 2025. Leveraging the collaborative power of AI and citizen science for sustainable development. *Nature Sustainability*, 8(2): 125–132.

Future of Life Institute. 2025. EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act. https://artificialintelligenceact.eu.

Groves, L.; Peppin, A.; Strait, A.; and Brennan, J. 2023. Going public: The role of public participation approaches in commercial AI labs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1162–1173. New York, NY, USA: ACM.

Hao, K.; and Seetharaman, D. 2023. Cleaning up ChatGPT takes heavy toll on human workers. *The Wall Street Journal*. https://www.wsj.com/tech/chatgpt-openai-content-abusive-sexually-explicit-harassment-kenya-workers-on-human-workers-cf191483.

Hartmann, D.; Laranjeira de Pereira, J. R.; Streitbörger, C.; and Berendt, B. 2025. Addressing the regulatory gap: Moving towards an EU AI audit ecosystem beyond the AI Act by including civil society. *AI & Ethics*, 5: 3617–3638.

Hirota, Y.; Andrews, J.; Zhao, D.; Papakyriakopoulos, O.; Modas, A.; Nakashima, Y.; and Xiang, A. 2024. Resampled datasets are not enough: Mitigating societal bias beyond single attributes. In *Proc. of the 2024 Conference on Empirical Methods in Natural Language Processing*, 8249–8267.

Hohendanner, M.; Ullstein, C.; Miyamoto, D.; Huffman, E. F.; Socher, G.; Grossklags, J.; and Osawa, H. 2024a. Metaverse perspectives from Japan: A participatory speculative design case study. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2).

Hohendanner, M.; Ullstein, C.; Onyekwelu, B. A.; Katirai, A.; Kuribayashi, J.; Babalola, O.; Ema, A.; and Grossklags, J. 2025. Initiating the Global AI Dialogues: Laypeople perspectives on the future role of genAI in society from Nigeria, Germany and Japan. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. New York, NY, USA: Assoc. for Computing Machinery.

Hohendanner, M.; Ullstein, C.; Socher, G.; and Grossklags, J. 2024b. "Good and scary at the same time" – Exploring citizens' perceptions of a prospective metaverse. *IEEE Pervasive Computing*, 23(1): 27–36.

IAP2. 1999. IAP2 Spectrum Evolution. https://www.iap2.org/page/SpectrumEvolution.

IAP2. 2025. IAP2 Public Participation Spectrum. https://iap2.org.au/resources/spectrum/.

ISO/IEC. 2023. ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. https://www.iso.org/standard/81230.html.

Kallina, E.; and Singh, J. 2024. Stakeholder involvement for responsible AI development: A process framework. In *Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '24. New York, NY, USA: ACM.

Kraft, P.; and Bansler, J. P. 1992. The collective resource approach: The Scandinavian experience. In *Proceedings of the Participatory Design Conference*, PDC '92, 127–135. Palo Alto, CA, USA: Computer Professionals for Social Resp.

Krommendijk, J.; and Borgesius, F. Z. 2023. How to read EU legislation? https://eulawanalysis.blogspot.com/p/how-to-read-eu-legislation.html?

Kulynych, B.; Madras, D.; Milli, S.; Raji, I. D.; Zhou, A.; and Zemel, R. 2020. Participatory Approaches to Machine Learning. Workshop Proceedings, ICML '20.

Lee, M. K.; Kusbit, D.; Kahng, A.; Kim, J. T.; Yuan, X.; Chan, A.; See, D.; Noothigattu, R.; Lee, S.; Psomas, A.; et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–35.

Mantelero, A. 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, 54: 1–18.

Martin Jr., D.; Prabhakaran, V.; Kuhlberg, J.; Smart, A.; and Isaac, W. S. 2020. Participatory Problem Formulation for Fairer Machine Learning Through Community Based System Dynamics. In *Machine Learning in Real Life, ICLR 2020 Workshop*.

Mökander, J.; Juneja, P.; Watson, D. S.; and Floridi, L. 2022. The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What can they learn from each other? *Minds and Machines*, 32(4): 751–758.

National Institute of Standards and Technology. 2023. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework.

Nekoto, W.; Marivate, V.; Matsila, T.; Fasubaa, T.; Kolawole, T.; Fagbohungbe, T.; Akinola, S. O.; Muhammad, S. H.; Kabongo, S.; Osei, S.; et al. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2144–2160.

Papakyriakopoulos, O.; and Mboya, A. M. 2023. Beyond algorithmic bias: A socio-computational interrogation of the Google search by image algorithm. *Social Science Computer Review*, 41(4): 1100–1125.

Pretty, J. N. 1995. Participatory learning for sustainable agriculture. *World Development*, 23(8): 1247–1263.

Publications Office of the European Union. 2022. *Interinstitutional style guide*. Publications Office of the European Union. https://data.europa.eu/doi/10.2830/215072.

Radiya-Dixit, E.; and Neff, G. 2023. A sociotechnical audit: Assessing police use of facial recognition. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1334–1346. New York, NY, USA: Association for Computing Machinery.

Rebrean, M.-L.; and Malgieri, G. 2025. Vulnerability in the EU AI Act: Building an interpretation. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 1985–1997. New York, NY, USA: Association for Computing Machinery.

Sharma, M.; Tong, M.; Mu, J.; Wei, J.; Kruthoff, J.; Goodfriend, S.; Ong, E.; Peng, A.; et al. 2025. Constitutional classifiers: Ddefending against universal jailbreaks across thousands of hours of red teaming. arXiv:2501.18837.

Sloane, M.; Moss, E.; Awomolo, O.; and Forlano, L. 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*, EAAMO '22. New York, NY, USA: ACM.

Sundblad, Y. 2010. UTOPIA: Participatory design from Scandinavia to the world. In Impagliazzo, J.; Lundin, P.; and Wangler, B., eds., *IFIP Conference on History of Nordic Computing*, 176–186. Berlin, Germany: Springer.

The Collective Intelligence Project. 2023. Participatory AI Risk Prioritization. https://www.cip.org/research/participatory-ai-risk-prioritization.

The White House. 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Executive Order 14110, Federal Register.

The White House. 2025. Removing Barriers to American Leadership in Artificial Intelligence. Executive Order 14179, Federal Register.

Ullstein, C.; Hohendanner, M.; and Grossklags, J. 2025. Anticipating risks and identifying governance measures for the use of genAI and FPT: Citizens' perspectives across six countries. In Weerts, H.; Pechenizkiy, M.; Allhutter, D.; Corrêa, A. M.; Grote, T.; and Liem, C., eds., *Proceedings of Fourth European Workshop on Algorithmic Fairness*, volume 294 of *PMLR*, 389–395.

Ullstein, C.; Hohendanner, M.; Sharma, N.; Ivanov, S. E.; Tsipov, G.; and Grossklags, J. 2025. Bridging perspectives for socially sustainable AI: What people across 12 countries think about genAI and FPT, and what policymakers want to know their opinions on. In *Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '25, 261–283. New York, NY, USA: Association for Computing Machinery.

Ullstein, C.; Pfeiffer, J. K.; Hohendanner, M.; and Grossklags, J. 2024. Mapping the stakeholder debate on facial recognition technologies: Review and stakeholder workshop. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, CSCW Companion '24, 429–436. New York, NY, USA: Association for Computing Machinery.

U.S. Congress. 2022. Algorithmic Accountability Act of 2022. https://www.congress.gov/bill/117th-congress/senate-bill/3572.

White, S. C. 1996. Depoliticising development: The uses and abuses of participation. *Development in Practice*, 6(1): 6–15.

Wilson, C. 2022. Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly*, 39(1): 1–10.

Wolf, C. T.; Zhu, H.; Bullard, J.; Lee, M. K.; and Brubaker, J. R. 2018. The changing contours of "participation" in data-driven, algorithmic ecosystems: Challenges, tactics, and an agenda. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18 Companion, 377–384. New York, NY, USA: Association for Computing Machinery.

Young, M.; Ehsan, U.; Singh, R.; Tafesse, E.; Gilman, M.; Harrington, C.; and Metcalf, J. 2024. Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday*, 29(4).

Zhang, A. Q.; Suh, J.; Gray, M. L.; and Shen, H. 2025. Effective automation to support the human infrastructure in AI red teaming. *Interactions*, 32(4): 58–61.

Zhao, D.; Andrews, J.; Papakyriakopoulos, O.; and Xiang, A. 2024. Measuring diversity in datasets. In *Workshop on Data-centric Machine Learning Research, ICLR 2024 Workshop*.