Why user involvement in AI?
And how to assess it.

Tuva Falk

AI Policy Lab
Umeå University, Sweden
id21tfk@cs.umu.se

Abstract. Effective artificial intelligence (AI) system design requires alignment with user expectations, yet the development process often prioritizes technical performance over human needs. Most AI systems optimize for efficiency, accuracy, and performance benchmarks, while overlooking fairness, transparency, usability, and cultural alignment. This narrow focus has led to well-documented failures, including biased hiring algorithms, discriminatory facial recognition systems, and opaque automated decision-making in critical areas such as credit scoring and policing [1, 2].
Without early stakeholder involvement, meaningful user evaluation, and the integration of community values, AI risks deepening societal inequities rather than mitigating them. To shift AI development from a purely technical paradigm to a human-first model, this policy brief proposes a three-pillar framework:

- **Early User Involvement**: AI systems must incorporate stakeholder participation from the outset to ensure diverse perspectives and prevent systemic biases from becoming embedded in models.
- **UX Evaluation Metrics**: AI performance should not be measured solely on accuracy and efficiency but should also incorporate usability, explainability, and user-driven evaluation criteria.
- **Value-Sensitive AI Design**: AI models should mirror the values of the communities in which they operate, enhancing system effectiveness and social acceptability.

Embedding these principles into AI governance can help ensure that AI development prioritizes human needs, social alignment, and public trust over purely technical efficiency.

1 Pillar 1: Early User Involvement - The Key to Fair AI

1.1 The Challenge

AI systems are often developed using historical data, which reflect existing biases. As a result, these systems risk perpetuating inequalities, reinforcing discrimination, and disproportionately impacting marginalized communities. While de-biasing tools and algorithmic fairness interventions exist, they are largely reactive, designed to correct biases after deployment rather than preventing them at the design stage [3]. The lack of early stakeholder involvement in AI development means technical optimization often takes precedence over human values like fairness and inclusivity, leading to AI systems that fail to account for real-world diversity and are prone to bias [2].

Furthermore, AI development is inherently sociotechnical, meaning that technical decisions are deeply intertwined with societal values, regulatory environments, and human behaviour [4]. AI Risk Management Frameworks recognize this interplay, emphasizing that AI risks, and benefits, emerge from the interaction between technological design choices and their societal contexts. However, to operationalize the positive effects of the sociotechnical system end users and impacted stakeholders need to be engaged early in the AI development lifecycle [5].

1.2 The Case for Early User Involvement

Integrating participatory design into AI development provides a structured approach to mitigating risks and maximizing benefits in sociotechnical systems. Liao et al. [4] highlight the critical role of UX methodologies in Responsible AI, emphasizing that early engagement with UX researchers helps surface stakeholder concerns. These insights inform decisions on AI model objectives, data selection, and evaluation criteria, fostering greater transparency and accountability. By incorporating input from diverse usersâespecially those most affected by algorithmic decisionsâparticipatory design ensures AI systems reflect a broader range of values and preferences. Research shows that user involvement enhances the interpretability and perceived fairness of AI outputs [6]. Early stakeholder participation enables proactive adjustments to AI objectives, training data, and evaluation metrics, promoting fairness and accountability throughout the development process.

Furthermore, UX and participatory design align closely with established AI governance frameworks such as the EU AI Act, which mandates human oversight and risk-based AI categorization [7]. Integrating user research and human-centred design processes into AI development provides a structured mechanism for operationalizing these governance principles, ensuring AI deployment aligns with both regulatory standards and societal expectations.

1.3 Policy Recommendations

1. **Encourage participatory AI design in high-risk applications**: AI system design should engage in stakeholder analysis, consultations, co-design workshops and iterative testing, ensuring systems are designed with real-world contexts in mind. before system design.
2. **Require stakeholder testing before AI deployment**: Introduce mandatory pre-deployment user testing as part of AI bias mitigation strategies. AI systems should be assessed not only through quantitative fairness metrics but also qualitative evaluations with diverse user groups, ensuring early identification of algorithmic blind spots and unintended harms.

3. **Implement user-first AI fairness audits**: Develop regulatory AI fairness audit frameworks that prioritize user experience and social impact, moving beyond a purely technical approach. AI audits should include human-centred criteria, such as explainability, accessibility, and contextual appropriateness, ensuring ethical and responsible AI deployment.
4. **Establish interdisciplinary teams**: Including UX professionals, ethicists, and domain experts to oversee AI development in sensitive areas and high risk areas.

2 Pillar 2: UX Evaluation Metrics - Establishing Accountability and Usability in AI

2.1 The Challenge

Traditional AI evaluation methods prioritize technical accuracy, efficiency, and computational performance, yet these factors alone do not determine an AI system's real-world effectiveness [8]. The absence of usability and UX metrics in AI assessment presents significant risks. Users often struggle to interpret AI-generated outputs [9], when users cannot understand the rationale behind an AI decision, they lack the ability to question or challenge unjust or discriminatory outcomes [10]. This opacity further exacerbates power imbalances between AI developers and users, as individuals affected by algorithmic decisions are often left without the knowledge, tools, or mechanisms to dispute an unfair result.

Current AI governance structures rely heavily on algorithmic audits and bias detection tools to assess AI fairness, but these methods often overlook the UX dimension of accountability and usability [6]. Evaluating AI based on statistical fairness ignores key human-centred concerns such as whether users understand AI recommendations, whether they trust the system, and whether the AI meets their expectations. Without structured UX-based evaluation frameworks, there is no standard way to measure whether AI is truly responsible [8].

2.2 The Case for UX Evaluation Metrics

User involvement in the selection of UX evaluation metrics offers a way to measure explainability, fairness, and user acceptance in a systematic manner. Research has shown that incorporating these metrics as early design materials can enhance alignment with user expectations and improves the overall usability of AI systems [11, 12]. Liao et al. [4] demonstrate that incorporating user feedback loops and participatory UX metrics significantly improves AI adoption, trust, and perceived fairness.

Zheng & Huang [11] emphasize that UX evaluation metrics should be integrated into participatory design processes from the outset, ensuring that AI developers do not impose predefined technical benchmarks that may misalign with actual user needs. Their study found that when users are actively involved in defining evaluation criteria, the resulting AI systems are more intuitive and better accepted by diverse user groups. Similarly, Bauer & Zangerle [8] argue that AI models must move beyond performance-based assessments, such as precision and recall, and aim to incorporate evaluation methodologies that assess whether the system provides value to end users. Without this shift, AI systems risk prioritizing technically optimal outcomes that fail to align with human-first priorities, ultimately reducing their real-world usability and effectiveness.

> Building on findings from our recent study, which surveyed 66 participants, we examined whether technical and non-technical stakeholders differ in their perspectives on AI evaluation metrics. While the results did not reveal significant differences in priority between the two groups, they did highlight that technical stakeholders had a more nuanced understanding of the evaluation metrics. Furthermore, the study examined the implications of correlations between design features of a recommender system, such as diversity and novelty in recommendations, and evaluation values, including perceived effectiveness and perceived usefulness. The observed correlations indicate that early incorporation of UX evaluation metrics helps streamline the design process, ensuring user preferences are aligned with the development of AI systems. Specifically, integrating evaluation values into participatory design enables a deeper understanding of user priorities, as demonstrated by strong correlations between features like quick, accurate updates and user-perceived effectiveness and usefulness. Identifying these associations allows designers to gain insights into the most critical aspects of AI system usability, ensuring that user needs and expectations are embedded from the early stages of development. This not only enhances the relevance of the design but also empowers users by ensuring their preferences actively shape AI development, ultimately fostering higher adoption and trust [12].

Moreover, UX evaluation metrics can serve as regulatory tools for AI governance. The EU AI Act has introduced risk-based AI categorization but currently lacks structured methods for assessing whether AI meets user-centric fairness and explainability criteria [7]. Establishing UX-driven evaluation methods would provide regulators with indicators of AI trustworthiness, making compliance assessments more transparent and enforceable [5].

2.3 Policy Recommendations

1. **Include User-Selected UX Evaluation Metrics in AI System Design**: AI system development should involve end users in the definition and selection of UX evaluation metrics as part of the early design process. This participatory approach ensures that AI systems are evaluated based on criteria that matter most to users, such as explainability, trust, and usability, rather than solely on technical performance metrics like accuracy and efficiency [11, 6, 4].

2. **Establish User-First Evaluation Metrics to Assess Usability and User Preferences**: Regulatory frameworks should require AI evaluation to incorporate usability-focused UX metrics, such as user satisfaction, explainability, fairness, and perceived effectiveness, rather than relying exclusively on technology-driven benchmarks like precision and recall [8].

3 Pillar 3: Value-Sensitive AI Design - Embedding Community Values into Algorithmic Decision-Making

3.1 The Challenge

Traditional AI development prioritizes efficiency, accuracy, and scalability, often assuming that algorithmic solutions are universally applicable across different contexts. However, this one-size-fits-all approach fails to capture the complexity of human values and community norms, leading to systems that may be inappropriate, misaligned, or even harmful when deployed in diverse communities [13, 14].

Failure to embed community values into AI models has resulted in algorithmic discrimination, cultural misalignment, and public distrust. Examples include, facial recognition biases where AI models trained on non-representative datasets fail to recognize individuals from certain demographic groups, leading to misidentifications and exclusion [2]. Automated decision-making in finance and hiring, where historical biases in training data result in unfair loan denials and discriminatory hiring outcomes [1]. AI-driven content moderation systems that struggle to understand context-specific norms, leading to unjust censorship or the failure to flag harmful content [15].

These failures underscore the need for Value-Sensitive Design methodologies, which integrate societal and community sensitive values directly into AI development. Without proactive efforts to align AI systems with community values, ethical principles, and cultural expectations, AI risks becoming a reinforcer of existing inequalities rather than a tool for progress.

3.2 The Case for Value-Sensitive AI Design

Value-Sensitive Design provides a structured approach to embedding ethical, social, and cultural considerations into AI systems. Unlike traditional algorithmic fairness approaches, which focus primarily on bias detection and mitigation after deployment, Value-Sensitive Design seeks to embed societal values into AI system architecture from the outset [16].

Zhu [14] argues that AI systems are more effective and better received when they mirror the values of the communities they serve. Their framework for Value-Sensitive Algorithm Design demonstrates that incorporating cultural, ethical, and community norms into AI development enhances the system's practical relevance and usability.

This principle is supported by empirical studies:

– **Participatory AI development increases trust**: Research shows that AI systems developed with community-driven value assessments have higher adoption rates and are perceived as more legitimate than those created without such input [6].
– **Context-aware AI improves decision-making**: AI models that integrate local knowledge, historical context, and cultural values are less likely to produce harmful unintended consequences [4].
– **Ethical AI regulation benefits from Value-Sensitive Design principles**: Governments that implement value-sensitive regulatory frameworks improve AI transparency and reduce legal and reputational risks for developers [13].

3.3 Policy Recommendations

1. **Require Value-Sensitive Impact Assessments Before Deployment**: Mandate that AI projects include structured value assessments to identify cultural, ethical, and community-specific considerations relevant to deployment contexts.
2. **Incentivize the Development of Context-Aware AI Models**: Provide funding and certification incentives for AI projects that demonstrate alignment with community values, increasing industry adoption of value-sensitive design.
3. **Create AI Requirements Based on Community Values**: Mandate AI transparency reporting, requiring AI systems to justify decisions in ways that align with public expectations and values.

4 Conclusion

Ensuring meaningful user involvement in AI development is not just an ethical imperative, it is a strategic necessity for creating transparent, accountable, and widely accepted AI systems. This brief has outlined the risks of a purely technical approach to AI design, highlighting how user-driven evaluation metrics can enhance fairness, usability, and trust. However, translating these insights into actionable policies requires overcoming practical challenges, including the feasibility of integrating diverse user perspectives, potential conflicts between stakeholders, and the perceived trade-off between efficiency and inclusivity.

To address these concerns, policymakers must take proactive steps to embed participatory design into AI governance frameworks. This requires establishing interdisciplinary oversight bodies, mandating stakeholder testing for high-risk applications, and redefining AI evaluation criteria to include usability, explainability, and fairness alongside traditional performance metrics. While implementation may require additional resources and infrastructure, the long-term benefits of improved adoption, reduced bias, and increased public trust, far outweigh the initial costs.

The time for action is now. Governments, industry leaders, and regulatory bodies must move beyond rhetoric and institutionalize mechanisms for user participation in AI design. By adopting a human-first approach, we can shift AI development from a predominantly technical exercise to a process that truly serves society's diverse needs.

References

[1] Shams, R.A., Zowghi, D., Bano, M.: AI and the quest for diversity and inclusion: a systematic literature review. AI and Ethics (November 2023)
[2] Aizenberg, E., Van Den Hoven, J.: Designing for human rights in AI. Big Data & Society **7**(2) (July 2020)
[3] Hall, P., Ellis, D.: A systematic review of socio-technical gender bias in AI algorithms. Online Information Review **47**(7) (November 2023) 1264–1279
[4] Liao, Q.V., Vorvoreanu, M., Subramonyam, H., Wilcox, L.: UX Matters: The Critical Role of UX in Responsible AI. interactions **31**(4) (June 2024) 22–27
[5] Hansen, N.B., Dindler, C., Halskov, K., Iversen, O.S., Bossen, C., Basballe, D.A., Schouten, B.: How Participatory Design Works: Mechanisms and Effects. In: Proceedings of the 31st Australian Conference on Human-Computer-Interaction, Fremantle WA Australia, ACM (December 2019) 30–41
[6] Delgado, F., Yang, S., Madaio, M., Yang, Q.: 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 (October 2023) 1–23
[7] : EU AI Act: first regulation on artificial intelligence (August 2023)
[8] Bauer, C., Zangerle, E., Said, A.: Exploring the Landscape of Recommender Systems Evaluation: Practices and Perspectives. ACM Transactions on Recommender Systems **2**(1) (March 2024) 1–31
[9] Yang, Q., Steinfeld, A., RosÃ©, C., Zimmerman, J.: Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, New York, NY, USA, Association for Computing Machinery (April 2020) 1–13
[10] Shneiderman, B.: Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. ACM Trans. Interact. Intell. Syst. **10**(4) (October 2020) 26:1–26:31
[11] Zheng, Q., Huang, Y.: "Begin with the End in Mind": Incorporating UX Evaluation Metrics into Design Materials of Participatory Design. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg Germany, ACM (April 2023) 1–7
[12] Falk, T.: Desining Responsible AI: The Power of User-Selected Metrics. In: Student Conference in Computing Science. (December 2024)
[13] Tacheva, J., Ramasubramanian, S.: AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. Big Data & Society **10**(2) (July 2023)
[14] Zhu, H., Yu, B., Halfaker, A., Terveen, L.: Value-Sensitive Algorithm Design: Method, Case Study, and Lessons. Proceedings of the ACM on Human-Computer Interaction **2**(CSCW) (November 2018) 1–23

[15] Bossen, C., D.C., Iversen, O.: Evaluation in participatory design: a literature survey. In: 14th Participatory Design Conference: Full papers. Volume 1., ACM (2016) 1–7
[16] Friedman, B., Kahn, P.H., Borning, A.: Value Sensitive Design: Theory and Methods