Negotiating AI Boundaries Through Participatory Futuring

Peer-reviewed author version

Negotiating AI Boundaries Through Participatory Futuring

JARNE THYS, UHasselt — Hasselt University, Digital Future Lab — Flanders Make, Belgium
MARISELA GUTIERREZ LOPEZ, Bristol Digital Futures Institute, University of Bristol, United Kingdom
EVA GEURTS, UHasselt — Hasselt University, Digital Future Lab — Flanders Make, Belgium
DAVY VANACKEN, UHasselt — Hasselt University, Digital Future Lab — Flanders Make, Belgium
GUSTAVO ROVELO RUIZ, UHasselt — Hasselt University, Digital Future Lab — Flanders Make, Belgium

The increasing deployment of AI systems in consequential domains has intensified calls to democratize AI governance. However, existing participatory approaches often involve stakeholders only after key system boundaries have already been set, and tend to frame disagreement as a problem to be resolved rather than a condition to be engaged. In this paper, we argue that many AI governance challenges are fundamentally boundary-setting problems, characterized by irreducible value conflicts over what AI should do, under what conditions, and who gets to decide. Drawing on agonistic pluralism and participatory design, we propose to apply a four-stage Participatory Boundary Negotiation (PBN) approach to consequential AI deployments. We demonstrate PBN's application through an illustrative use case of AI-assisted grading in universities, showing how the method grounds deliberation in situated practices, surfaces conflicts through speculative futures, diagnoses their underlying conflict types, and supports negotiation through boundary objects. Rather than seeking consensus, this approach treats disagreement as a legitimate and productive feature of democratic AI governance, thereby contributing to participatory futuring research in consequential contexts.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**; **Participatory design**.

Additional Key Words and Phrases: Participatory Design, Participatory Futures, AI Boundary Negotiation, Responsible AI, Democratization of AI

1 Framing AI Boundary-Setting in Participatory Design

The growing deployment of AI across consequential domains has led to calls to democratize AI governance [14]. However, a survey of 80 participatory AI projects reveals significant limitations: 94% of stakeholder selection remained under the control of project teams, 85% of participation occurred only after the system scope was determined, and just 5% allowed stakeholders to rule out AI as a solution [3]. This gap is particularly evident in boundary-setting problems, where objectives, limits, and decision authority are determined top-down despite conflicting values [8].

Existing approaches struggle to productively engage with value conflicts in AI governance. Value Sensitive Design (VSD) offers systematic methods for integrating stakeholder values [6], but assumes conflicts can be reconciled through better requirements and iterative design. This is insufficient for fundamental value conflicts: disagreements rooted in incompatible views that cannot be resolved through rational deliberation alone [2]. These challenges point out the need for approaches grounded in agonistic pluralism, which recognize that some conflicts are irreducible and that legitimate adversaries can contest boundaries without seeking to eliminate difference [4, 10]. Rather than treating disagreement as friction to be designed away, agonistic approaches position contestation as a necessary feature of democratic technology governance.

We propose to use *Participatory Boundary Negotiation* (PBN) as a method to integrate agonistic pluralism into AI boundary-setting. Drawing on speculative enactment, value conflict typologies,

Authors' Contact Information: Jarne Thys, jarne.thys@uhasselt.be, UHasselt — Hasselt University, Digital Future Lab — Flanders Make , Diepenbeek, Belgium; Marisela Gutierrez Lopez, marisela.gutierrezlopez@bristol.ac.uk, Bristol Digital Futures Institute, University of Bristol, Bristol, United Kingdom; Eva Geurts, eva.geurts@uhasselt.be, UHasselt — Hasselt University, Digital Future Lab — Flanders Make , Diepenbeek, Belgium; Davy Vanacken, davy.vanacken@uhasselt.be, UHasselt — Hasselt University, Digital Future Lab — Flanders Make , Diepenbeek, Belgium; Gustavo Rovelo Ruiz, gustavo.roveloruiz@uhasselt.be, UHasselt — Hasselt University, Digital Future Lab — Flanders Make , Diepenbeek, Belgium.

and boundary objects, PBN treats contested boundaries as sites of legitimate ongoing contestation. Our four-stage method uses alternative futures to surface conflicts, categorizes disagreements as resolvable, constitutive, or power-laden, and employs task cards as boundary objects to enable negotiation across incompatible views while producing actionable boundary sets.

2 Participatory Boundary Negotiation: A four-Stage Method

In this section, we propose Participatory Boundary Negotiation as a method for contexts in which stakeholders disagree about what AI should be allowed to do, under what conditions, and who should decide. Rather than treating disagreement as friction to resolve, PBN makes contested boundaries observable and negotiable through four interconnected stages.

2.1 Stage 0: Grounding in Situated Contexts

Before introducing speculative scenarios, PBN requires understanding the specific practices, histories, and power relations that shape how stakeholders experience AI in their contexts. Drawing on participatory design's emphasis on mutual learning [11], this stage involves stakeholders in identifying which aspects of their current situations matter most for boundary-setting. Rather than treating stakeholders as validators of researcher-designed scenarios [3], Stage 0 positions them as experts who co-create the provocations guiding subsequent stages.

This grounding can, for example, involve participatory mapping of stakeholder ecosystems and authority relations [7], or collaborative documentation of past experiences with technology governance. For example, journalists' experiences with AI reveal context-specific value tensions [9] that generic scenarios would miss. Stage 0 ensures that speculative futures in Stage 1 resonate with stakeholders' lived realities rather than imposing external framings, while also surfacing existing power asymmetries [7] and varying technical literacies [17] that will shape how conflicts emerge and are negotiated in subsequent stages.

2.2 Stage 1: Eliciting Conflicts Through Speculative Scenarios

Participatory AI projects often begin with an AI-first framing and predefined specifications, so questions about whether AI should be deployed may fall outside the scope of participation [3]. Abstract value discussions may fail to surface situated conflicts that matter to stakeholders.

We propose speculative enactment [5] using Dator's four alternative futures [1], which we map on GenAI use: Continued Growth (minimal boundaries, efficiency-focused), Collapse (absolute restrictions after incidents), Discipline (top-down regulation), and Transformation (context-dependent plural boundaries). These scenarios function as provocations, not predictions. By introducing extreme but plausible futures through design fiction, role-play, or interactive prototypes, stakeholders can articulate fears and red lines that might otherwise remain implicit. These surfaced conflicts then require diagnosis (Stage 2) to determine whether they stem from misaligned priorities, incompatible worldviews, or power asymmetries.

2.3 Stage 2: Diagnosing Conflict Types

VSD [6] methods often assume conflicts can be reconciled through better requirements, but offer no method for determining whether deliberation, compromise, or recognition of irreconcilable differences is appropriate. Without diagnosing the causes of conflicts, facilitators risk false consensus or unproductive deadlock. Prior research shows the value of understanding the kinds of value tensions before trying to resolve them. For example, studies of AI-supported journalism find that the same AI systems may be seen as both supporting and undermining core professional values, such as truth, impartiality, public interest, and originality, depending on context and perspective [9].

To move beyond treating all disagreements as design problems to be resolved, we adapt Schön and Rein's typology of policy framing conflicts [13] to diagnose three conflict types: (1) Resolvable conflicts: shared frames, different priorities (addressable through VSD) [13]; (2) Constitutive frame conflicts: incompatible worldviews requiring agonistic negotiation and provisional agreements [10, 13]; (3) Authority conflicts: disputes over whose interpretation counts, requiring explicit power arrangements [7, 13]. This diagnosis directly informs Stage 3's negotiation strategy: resolvable conflicts through technical solutions, constitutive conflicts through agonistic negotiation, or authority conflicts through power restructuring.

2.4 Stage 3: Observing Negotiation Through Co-Design

Participatory projects typically position stakeholders as validators of predefined options [3], lacking mechanisms to translate values into concrete governance decisions.

We propose collaborative card sorting [11, 15] where mixed stakeholder groups categorize AI use cases into green (permitted), yellow (conditional), and red (prohibited). Task cards function as boundary objects [16], enabling coordination across incompatible worldviews by forcing stakeholders to specify conditions under which contested applications become acceptable. The output documents both agreements and persistent contestation, producing a negotiated governance artifact that deliberately resists technical closure. In line with previous HCI futuring research [12], we treat AI boundaries as provisional and contested rather than as problems awaiting definitive resolution.

3 Illustrative use case and Discussion

To illustrate PBN's potential application, we present a hypothetical case of AI-assisted grading in universities. In Stage 0, stakeholders (e.g., students, instructors, administrators) map current assessment practices and authority structures, revealing tensions between efficiency pressures and pedagogical values. Stage 1 presents four futures: Continued Growth (AI grades all assignments), Collapse (complete ban after bias incidents), Discipline (institutional review boards approve each use), and Transformation (departments set context-specific boundaries). Faculty express fears about deskilling; students worry about fairness; administrators prioritize scalability.

Stage 2 diagnoses conflicts: resolvable (e.g., AI for objective vs. subjective grading), constitutive (whether assessment should develop judgment vs. measure competencies), and authority (who decides pedagogical boundaries). Stage 3 uses card sorting where mixed groups categorize use cases: "AI-graded multiple choice" (green), "AI-suggested essay feedback, instructor reviews" (yellow with conditions: human override, transparency to students), "AI-generated final grades" (red for some, yellow for others; documented disagreement).

As this example demonstrates, PBN does not eliminate conflict but makes it productive. The method still has limitations: it cannot guarantee equitable outcomes when power asymmetries persist, requires organizational willingness to treat AI deployment as contestable, and produces provisional rather than final boundaries. Rather than viewing this instability as a methodological weakness, we argue it reflects participatory futuring's core challenge: foregrounding power, pluralism, and irreducible value conflict to make contested AI boundaries visible and negotiable rather than resolved away.

Acknowledgments

This work was supported by the Special Research Fund (BOF) of Hasselt University (BOF24OWB28).

References

[1] Jim Dator. 2019. Alternative Futures at the Manoa School. In *Jim Dator: A Noticer in Time: Selected work, 1967-2018*, Jim Dator (Ed.). Springer International Publishing, Cham, 37–54. doi:10.1007/978-3-030-17387-6_5

[2] Karl de Fine Licht. 2025. Resolving value conflicts in public AI governance: A procedural justice framework. *Government Information Quarterly* 42, 2 (June 2025), 102033. doi:10.1016/j.giq.2025.102033

[3] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3617694.3623261

[4] Carl DiSalvo. 2022. *Design as Democratic Inquiry: Putting Experimental Civics into Practice*. The MIT Press. doi:10.7551/mitpress/13372.001.0001

[5] Chris Elsden, David Chatting, Abigail C. Durrant, Andrew Garbett, Bettina Nissen, John Vines, and David S. Kirk. 2017. On Speculative Enactments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5386–5399. doi:10.1145/3025453.3025503

[6] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. In *Early engagement and new technologies: Opening up the laboratory*, Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman (Eds.). Springer Netherlands, Dordrecht, 55–95. doi:10.1007/978-94-007-7844-3_4

[7] John Gaventa. 2006. Finding the Spaces for Change: A Power Analysis. *IDS Bulletin* 37, 6 (2006), 23–33. doi:10.1111/j.1759-5436.2006.tb00320.x

[8] Emma Kallina, Thomas Bohné, and Jatinder Singh. 2025. Stakeholder Participation for Responsible AI Development: Disconnects Between Guidance and Current Practice. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 1060–1079. doi:10.1145/3715275.3732069

[9] Tomoko Komatsu, Marisela Gutierrez Lopez, Stephann Makri, Colin Porlezza, Glenda Cooper, Andrew MacFarlane, and Sondess Missaoui. 2020. AI should embody our values: Investigating journalistic values to inform AI technology design. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3419249.3420105

[10] Chantal Mouffe. 2013. *Agonistics: Thinking The World Politically*. Verso Books.

[11] Michael J. Muller. 2002. Participatory design: the third space in HCI. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. L. Erlbaum Associates Inc., USA, 1051–1068.

[12] Camilo Sanchez, Sui Wang, Kaisa Savolainen, Felix Anand Epp, and Antti Salovaara. 2025. Let's Talk Futures: A Literature Review of HCI's Future Orientation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–36. doi:10.1145/3706598.3713759

[13] Donald Schön and Martin Rein. 1994. Frame reflection: Toward the resolution of intractable policy controversies. *Basic Book* (1994).

[14] Elizabeth Seger, Aviv Ovadya, Divya Siddarth, Ben Garfinkel, and Allan Dafoe. 2023. Democratising AI: Multiple Meanings, Goals, and Methods. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. Association for Computing Machinery, New York, NY, USA, 715–722. doi:10.1145/3600211.3604693

[15] Donna Spencer. 2009. *Card sorting: Designing usable categories*. Rosenfeld Media.

[16] Susan Leigh Star and James R. Griesemer. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 3 (Aug. 1989), 387–420. doi:10.1177/030631289019003001

[17] Jarne Thys, Davy Vanacken, and Gustavo Rovelo Ruiz. 2025. Engineering Trustworthy Automation: Design Principles and Evaluation for AutoML Tools for Novices. doi:10.48550/ARXIV.2511.22352