Designing with Uncertainty:

LLM Interfaces as Transitional Spaces for Democratic Revival

Sylvie Delacroix[1]

King's College London & visiting professor at Tohoku University

Forthcoming in *Minds and Machines* (late 2025 / early 2026)

Abstract:

The integration of large language models (LLMs) into our conversational infrastructure presents a critical inflection point for democratic practice. While contemporary digital platforms systematically erode transitional conversational spaces-interfaces between private intuition and public deliberation where tentative thoughts can be explored-this paper argues that specialized LLM interfaces could potentially reconstruct these essential democratic environments. I propose a design framework for 'transitional conversational spaces' that leverages uncertainty expression not merely to prevent unwarranted epistemic confidence but to create communicative environments conducive to democratic capability development.

Drawing on theories of democratic deliberation and moral perception, this paper distinguishes between epistemic uncertainty (addressable through additional information) and hermeneutic uncertainty (concerning the inherently contestable nature of interpretation). The proposed framework emphasizes 'ensemble interfaces' that make visible the contingent nature of value judgments by presenting outputs from multiple models trained on different datasets.

The design principles outlined challenge tokenistic participation by advocating for substantive participatory infrastructure with features like 'tinkerability' (enabling communities to experiment with system configurations) and mechanisms that counter designer-centric development models. These principles stand in contrast to conventional 'participatory AI' approaches that treat engagement as merely instrumental to system optimisation rather than as constitutive of democratic practice.

This paper does not claim to solve all challenges of democratic participation but rather identifies one valuable design direction that could potentially enhance our collective capacity for exploratory dialogue. Implementation would require institutional transformations that align technological development with democratic values beyond current procedural approaches to AI governance.

[1] My sincere thanks go to Dartmouth College's Montgomery fellowship program, which afforded me crucial time for reflection on the core ideas presented in this article and the chance to discuss them in a public lecture. I am particularly thankful for the thought-provoking exchanges and useful suggestions provided by Jacopo Domenicucci. The astute questions raised during my research seminar at the Oxford Institute for Ethics in AI proved highly valuable, as were my discussions with Diana Robinson and my academic colleagues Massimo Renzo and Lorenzo Zucca.

1. Introduction

Recent advances in large language models (LLMs) have delivered impressive gains in fluency and coherence of automatically generated text (Bommasani et al., 2021). This ability to produce naturalistic and contextually appropriate responses intersects with our ingrained propensity to engage in the collaborative meaning-making processes inherent to conversation. The conjunction of these two forces-the novel ability to simulate conversational turn-taking and our drive to make sense of the world through dialogue-has catalysed a situation where we often find ourselves treating and experiencing these tools as conversational partners. It is this conversational experience-and not some inherent attribute (whether it be some putative intelligence, intentionality or otherwise)-that distinguishes LLMs from other cultural technologies.

If, despite their artificial, disembodied nature, LLMs' strikingly coherent replies lead us to extend the same 'interpretive charities' as those that lubricate ordinary discursive interactions with fellow humans, the difficulties we face transcend mere individual epistemic calibration concerns. These calibration issues have long been familiar to human-computer interaction experts, arising with any technological system presenting an opaque epistemic façade. Like other machine learning tools, LLMs exhibit this opacity while additionally inviting users into conversational dynamics that implicitly solicit interpretive trust. Significant research effort is currently devoted to finding ways for LLMs to communicate their outputs with some measure of reliability, in a bid to prevent unwarranted epistemic confidence. Important as it is, this line of work risks concealing other, no less significant implications.

The problem extends beyond merely identifying resources necessary for tempering the charitable stance we tend to bring to any conversation. More fundamentally, it stems from the reciprocal nature inherent in conversational exchange. Just like any conversational partner, LLMs will actively shape and condition our perceptual and interpretive habits. This co-evolutionary dynamic transcends individual interactions, operating at a systemic level. Given the scale of their deployment, one must consider LLMs' pervasive impact upon the normative practices within which they are embedded.

Many of these practices (from healthcare to education via most forms of democratic discourse) presuppose our ongoing capability to question and deliberate about the values embedded in these practices. This capability cannot be taken for granted. The decline in democratic discourse is often attributed to polarization and misinformation, yet these phenomena constitute symptoms rather than root causes of our current predicament. This paper critically considers the erosion of 'transitional conversational spaces'. At their best, such spaces provide mediating environments where tentative moral perceptions can develop across different registers of articulation without premature closure into fixed deliberative positions.

The erosion of these spaces is not accidental. It is built into the design of contemporary digital platforms. Social media environments systematically reward certainty and punish nuance, creating communicative ecosystems that privilege expressions of certainty over expressions of uncertainty, resolution over exploration, and conflict over collaboration. These

design choices reinforce conditions that undermine the ability to consider multiple perspectives: a capacity essential for democratic deliberation.

Specific design choices will determine whether particular LLM implementations ultimately exacerbate these concerning trends or offer potential alternatives in our communicative landscape. This paper considers specifically designed LLM interfaces that could function as a novel kind of 'transitional conversational spaces'. Drawing on Winnicott's (1971) concept of transitional objects, I argue that such specialized LLM interfaces could scaffold the development of the pre-reflective skills that underlie robust democratic practices. Importantly, this paper does not advocate that all LLM deployments should adopt this framework, but rather that this represents one valuable design direction that warrants serious exploration alongside other approaches to LLM development.

The paper proceeds in several stages. Section 2 establishes the relationship between uncertainty, moral perception, and democratic practice, highlighting how contemporary communicative environments systematically undermine the conditions necessary for productive sense-making practices. Section 3 examines the potential of specifically designed LLMs to function as transitional conversational spaces. The analysis focuses on interfaces' capacity to scaffold the recursive processes through which moral understanding develops across the spectrum of articulation—from tentative exploration to more formalised democratic engagement. Section 4 analyses the role of uncertainty in creating such transitional spaces, distinguishing between epistemic and hermeneutic forms of uncertainty and their distinct roles in democratic practice. Section 5 outlines principles for the design of LLM interfaces that yield such spaces, emphasizing the importance of 'ensembles' that make visible the contingent and contestable nature of political judgment. Finally, Section 6 considers the implications of this reconceptualization for democratic renewal and explores pathways for implementation.

2. The Erosion of transitional conversational spaces

2.1 Democracy and the Spectrum of Articulation

Democratic practice depends not only on formal deliberative processes but on the ongoing refinement of the perceptual and cognitive habits that allow citizens to recognize and respond to morally salient features of collective life. This refinement occurs across a 'spectrum of articulation'[2]-from tentative, exploratory exchanges to more formalized expressions suitable for public discourse (Williams 1981). Traditional accounts of democratic deliberation have often privileged the more articulate end of this spectrum, focusing on formal deliberative

[2] 'The public order, if it is to carry conviction, and also not to flatten human experience, has to find ways in which it can be adequately related to private sentiment, which remains more 'intuitive' and open to conflict than public rules can be. For the intuitive condition is not only a state which private understanding can live with, but a state which it must have as part of its life, if that life is going to have any density or conviction and succeed in being that worthwhile kind of life which human beings lack unless they feel more than they can say, and grasp more than they can explain' (Williams, 1981, p. 82).

processes while overlooking the vital work that unfolds within less structured communicative environments.

The bidirectional movement between different points on this spectrum requires the cultivation of intermediary discursive environments. Faith-based communities, civic associations, labour unions, and other forms of community organization have historically provided spaces[3] where citizens could refine their moral perceptions in dialogue with others before engaging in formal deliberative contexts. The erosion of these mediating institutions has created a widening gap between private intuition and public deliberation-a gap that contemporary digital architectures have exacerbated rather than bridged.

The resultant communicative ecosystem lacks 'potential space'[4]: a liminal domain of intersubjective experience where individual psychological interiority engages with collective symbolic frameworks through exploratory exchange (Winnicott, 1971). Without such transitional domains, citizens are propelled prematurely into polarized deliberative contexts where performative certainty becomes prerequisite for participation, effectively foreclosing the attentive exploration necessary for moral inquiry.

2.2 The Transformation of Democratic Capacity

The erosion of transitional conversational spaces undermines democratic capacity by compromising 'reflective disclosure'-the cognitive-dialogical process through which citizens recursively engage with multiple interpretive frameworks without immediate reduction to predetermined positions (Kompridis 2011). Absent such liminal spaces, the democratic ecosystem increasingly lacks the socio-cognitive infrastructure for 'deliberative reciprocity': the practice of acknowledging the legitimacy of diverse normative orientations while simultaneously advancing one's own (Gutmann and Thompson, 1996).

The manifestation of these democratic deficits occurs through complex reconfigurations of political discourse ecologies. While early digital critics like Pariser (2011) emphasized 'filter bubbles' as primary mechanisms of epistemic isolation, contemporary empirical investigations have substantially complicated this narrative. Bruns (2019) challenges the filter bubble thesis by documenting the substantial cross-ideological exposure that characterizes actual digital media consumption patterns. Similarly, Guess et al. (2018) demonstrate that online media ecosystems often feature more ideological diversity than offline counterparts, suggesting that digital fragmentation has been significantly overstated in popular discourse.

More nuanced analyses reveal a paradoxical relationship between exposure and polarization. Bail et al. (2018) found that exposing participants to opposing political perspectives increased rather than decreased polarization, particularly among conservative participants. These

[3] 'The norms and networks of civic engagement also powerfully affect the performance of representative government. That, at least, was the central conclusion of my study of the twenty regional governments instituted in Italy in 1970. Social capital, as embodied in horizontal networks of civic engagement, bolstered the performance of the polity and the economy...' (Putnam, 2000, p. 346).

[4] 'It is in the space between inner and outer world, which is also the space between people--the transitional space--that intimate relationships and creativity occur.' (Winnicott, 1971, p. 2).

findings suggest ‘motivated exposure’-selective engagement with diverse content that serves to reinforce rather than challenge existing beliefs (Dubois and Blank, 2018).

This complex landscape is characterized not by simple ideological sorting but by transformed modes of encountering alterity. What emerges is not merely informational isolation but diminished capacity for ‘agonistic pluralism’[5]: the ability to engage productively with normative difference (Mouffe, 2005). Democratic vitality depends not on eliminating contestation but on the ‘civilizing’ of conflict-establishing institutional architectures that facilitate engagement with difference in ways that generate intersubjective understanding rather than antagonistic entrenchment (Schudson, 1997).

2.3 Technological Architecture and Democratic Renewal

Contemporary digital platforms systematically reinforce patterns of communication that privilege certainty over uncertainty, resolution over exploration, and conflict over collaboration. The algorithmic architecture of these platforms prioritizes content that elicits strong emotional responses (Brady et al., 2017), creating communicative environments hostile to the tentative articulations necessary for moral and political development. This systematic devaluation of exploratory dialogue occurs across the digital ecosystem, from social media platforms to search engines and recommendation systems that reflect and reinforce existing patterns of attention allocation (Noble, 2018).

Creating new transitional conversational spaces presents significant challenges within this technological environment. Traditional approaches have focused on platform reform through regulation or alternative business models. These approaches, while important, may prove insufficient given the deeply embedded nature of contemporary attention economies. An alternative approach focuses on developing specific forms of technological mediation designed to support the conditions necessary for productive moral deliberation.

This is where specialized LLM interfaces present a potential opportunity. By constructing ‘attention-conscious architectures’ that privilege exploratory over resolutive dialogue, such interfaces could serve as new forms of transitional spaces. Such spaces would create the environmental conditions necessary for scaffolding the pre-reflective skills that underlie robust democratic practices in an increasingly mediated world.

3. LLMs as transitional conversational spaces

3.1 Technological Mediation and Deliberative Practice

The erosion of traditional mediating institutions necessitates innovative approaches to reconstructing the infrastructure essential for democratic deliberation. A specific subset of LLM interfaces-designed with deliberate attention to their mediational qualities-could

[5] ‘The creation of a democratic society involves the recognition that any social objectivity is ultimately political and has to show the traces of the exclusion which governs its constitution... A pluralist democracy makes room for the expression of dissent and conflicting interests and values.’ (Mouffe, 2005, p. 30).

establish environments that facilitate the interpretive capacities fundamental to democratic participation.

The design framework for transitional conversational spaces transcends instrumental perspectives that position language technologies as mere information retrieval mechanisms or task-completion utilities. This framework conceptualizes these specialized LLM interfaces as mediating technologies[6] that shape the relationship between humans and their interpretive engagement with the world (Ihde 1990). Rather than focusing on technical parameters such as accuracy or alignment, this perspective prioritizes the relational affordances that emerge through technological interaction.

Contemporary AI development paradigms predominantly emphasize algorithmic optimization, whereas transitional conversational spaces would instead prioritize capability enhancement. This capability-oriented approach to technological design recognizes how socio-technical systems can expand substantive freedoms and human agency. The capability approach ‘provides a compelling normative framework for the design of technologies’ by focusing on ‘human flourishing and the substantive freedom of individuals to achieve valuable 'beings and doings' (Oosterlaken, 2015). Consequently, LLM interfaces designed as transitional spaces would aim not merely to optimize performance metrics but to support human developmental capacities.

The distinctive contribution of transitional conversational spaces emerges from their orientation toward democratic capability-building. Unlike contemporary algorithmic systems that prioritize informational transmission or behavioural prediction, these specialized interfaces would measure their efficacy through their capacity to support the intersubjective negotiation of meaning across diverse interpretive frameworks. This shift aligns with ‘reflexive public deliberation’[7], as a set of processes that enable participants to ‘recursively examine and transform the conditions of their own discourse’ (Bohman, 2004).

3.2 Scaffolding Interpretive Capacities Through Dialogue

The development of moral perception-the ability to recognize morally salient features of situations-depends on dialogic engagement. This capacity for attention develops through conversation with others who may perceive different aspects of the same situation. The specialized LLM interfaces proposed here could potentially scaffold the development of interpretive capacities in several ways that transcend the limitations of existing digital communication environments.

First, transitional conversational interfaces could facilitate perspective-taking by presenting multiple interpretive frameworks for considering moral and political questions. Unlike search

[6] ‘Technologies transform our experience of the world and our perceptions and interpretations of our world, and we in turn become transformed in this process. Phenomenology interprets this mutual constitution of perceptions, embodied knowledge, technological artifacts, and social practices.’ (Ihde, 1990, p. 49).
[7] ‘Reflexivity in deliberation is required if citizens are to see their deliberative activities not merely as solving first-order problems but also as higher-order reflection upon the adequacy of the ways in which they frame, discuss, and solve these problems. Citizens address not only the specific topics of deliberation but also the procedures, rules and parameters of deliberation itself.’ (Bohman, 2004, p. 134).

technologies that aggregate information without contextual integration, these specialized interfaces could engage users in substantive dialogue about how different normative perspectives might interpret the same situation. This deliberative process could help develop ‘communicative democracy’-a conception of democratic practicethat emphasizes ‘the importance of narrative and situated knowledge in political judgment’ (Young, 2000).

Second, interfaces designed as transitional spaces could support the development of emerging moral perceptions that are not yet fully formed or coherent. By engaging with users' tentative expressions and helping to refine them, these systems could support the movement from ‘imperfect rationalisations’ to more fully articulated moral positions (Williams 1981). This would require systems designed to value exploratory thinking rather than premature resolution-a significant departure from prevalent communicative architectures.

Third, transitional conversational interfaces could model the interpretive virtues necessary for productive moral deliberation: openness to revision, engagement with counterarguments, and respect for different perspectives. By embodying these virtues in dialogue, such systems could help users develop their own deliberative capacities, creating a liminal space for the development of skills essential to democratic participation.

3.3 Bridging private intuition and Public Deliberation

Contemporary communicative environments often propel tentative thoughts immediately into highly polarized public contexts, lacking the intermediary spaces necessary for the development and refinement of inchoate moral perceptions.

This bridging function depends on maintaining a state of ‘productive uncertainty’-a condition in which dialogic exploration is prioritized over premature resolution. This form of uncertainty differs fundamentally from simple epistemic indeterminacy. It represents a deliberate suspension of judgment that creates space for the exploration of multiple interpretive possibilities. ‘Epistemic friction’[8] between diverse perspectives generates a productive tension that enhances rather than diminishes collective intelligence (Medina, 2013).

The capacity of transitional conversational interfaces to maintain productive uncertainty depends not primarily on quantitative measures of confidence but on qualitative features. Systems designed to privilege exploratory thinking would need to embody conversational virtues: responsiveness to tentative views, openness to revision, comfort with ambiguity, and patience with the meandering character of moral inquiry.

These specialised interfaces may be seen as addressing the need for ‘subaltern counterpublics’[9]-social spaces that combine elements of both exploratory and more

[8] ‘Epistemic friction’ occurs when ‘different ways of seeing things are brought into contact with one another in such a way that they can interrogate each other, challenge each other, and... stimulate the formation of new perspectives.’ (Medina, 2013, p. 50).

[9] ‘Subaltern counterpublics function as spaces of withdrawal and regroupment... [and] also function as bases and training grounds for agitational activities directed toward wider publics. It is precisely in the dialectic between these two functions that their emancipatory potential resides.’ (Fraser, 1990, p. 68).

formalized discourse (Fraser, 1990). Professional communities often serve this function, as do various forms of civil society organization. From that perspective, specialized LLM interfaces could serve as new forms of mediating institutions, providing the spaces necessary for the renewal of democratic capacities.

4. Uncertainty and Democratic Practice

4.1 Epistemic Versus Hermeneutic Uncertainty

To understand how specifically designed LLM interfaces might support democratic deliberation, one needs to distinguish between two different forms of uncertainty: epistemic and hermeneutic. Epistemic uncertainty concerns gaps in knowledge that could, in principle, be filled through additional information or improved methods. Hermeneutic uncertainty, by contrast, concerns the inherently open and contestable nature of interpretation itself-the fact that the same situation can be legitimately interpreted in multiple ways.

This distinction proves crucial for democratic contexts. Reasonable disagreement in democratic societies reflects not merely epistemic limitations but 'the burdens of judgment'-the fact that reasonable citizens with access to the same information may reach different conclusions due to differences in how they interpret that information (Rawls, 2005). The normative pluralism characteristic of democratic societies cannot be resolved through epistemic means alone; it requires ongoing intersubjective negotiation about the meanings of foundational values.

Current work on getting LLMs to communicate their outputs with some measure of reliability focuses primarily on epistemic dimensions. Methods such as calibrated confidence scores, verbal expressions of uncertainty, and semantic uncertainty estimation help users gauge the reliability of factual claims and identify potential hallucinations (Kadavath et al., 2022).[10] While these approaches can offer valuable assistance, they do not address the hermeneutic dimensions of uncertainty.

Rather than merely quantifying reliability, what democratic contexts require are conversational environments that keep interpretive possibilities open and invite collaborative exploration. Such environments would acknowledge 'the contest of interpretation'-the fact that the meaning of political principles and values is constantly being renegotiated through democratic practice (Benhabib, 1992).

4.2 Uncertainty Expression and the Cultivation of Moral Perception

The relationship between how uncertainty is allowed to surface in communicative environments and the development of moral perception remains remarkably under-theorized in both technological and democratic discourse. While extensive scholarly attention has focused on calibrated expressions of epistemic uncertainty, far less consideration has been

[10] The significance -and limits- of this work is explored further in (Delacroix et al., 2025).

given to how different modalities of uncertainty expression might enable or constrain the perceptual capacities that underlie moral understanding.

Moral perception-the cultivated ability to recognize ethically salient features within complex situations-emerges not primarily through abstract reasoning but through attentive engagement with particular cases (van Domselaar 2022) or persons (Van Grunsven, 2022). This perceptual capacity involves ‘a cultivated perceptual sensitivity to recognize morally salient features of situations, particularly how one's actions might affect others’ (Walker, 2007). Such attentiveness requires a willingness to suspend immediate judgment in favour of careful exploration-a stance that depends crucially on communicative environments that privilege inquiry over premature resolution.

The cultivation of this attentive exploration demands ‘other-responsiveness’: a receptivity to perspectives and experiences that differ substantively from one's own established frameworks.[11] This attentional practice can be understood as a form of ‘world-traveling’, where ‘the shift from being one person to being a different person... requires that we be willing to be vulnerable, to be more than minimally open to surprises, to self-construction or reconstruction’ (Lugones, 1987). This form of world-traveling represents not merely an epistemological orientation but an attentional practice with significant implications for moral perception.

The expression of uncertainty plays a pivotal role in establishing communicative conditions conducive to such attentional practices. When conversational partners express uncertainty, they signal openness to collaborative exploration rather than commitment to predetermined conclusions. This communicative stance creates intersubjective space for tentative moral insights-particularly those that may challenge established normative frameworks.[12]

For specialized LLM interfaces designed as transitional conversational spaces, this raises distinctive design challenges. If uncertainty is expressed merely as quantified confidence levels for discrete factual claims, its potential to support moral perception remains severely constrained. If, however, uncertainty is expressed as an invitation to collaborative exploration of interpretive possibilities, it can potentially enhance the development of moral perception by creating legitimate spaces for tentative articulation. This distinction parallels the differentiation between ‘thin’ and ‘thick’ forms of deliberation-where thin deliberation focuses primarily on aggregating preferences, while thick deliberation centers on transformative dialogue that opens participants to new perspectives and possibilities.

This analysis suggests that transitional conversational spaces must be designed to express uncertainty not merely as an epistemic limitation to be minimized, but as a generative condition that enables new forms of moral perception to emerge. The expression of

[11] The role played by this ‘other-responsiveness’ in the extent to which our habits remain at the service of our ethical life is further developed in (Delacroix 2022a).
[12] Murdoch formulates this in terms of ‘a just and loving gaze directed upon an individual reality... the extremely difficult realization that something other than oneself is real’ (Murdoch, 1970).

uncertainty in these specialized interfaces would thus serve not primarily to calibrate epistemic confidence but to create the conditions for transformative dialogue

4.3 Uncertainty Expression and the Cultivation of Political Judgment

The implications of uncertainty expression extend beyond individual moral perception to encompass collective processes of political judgment. Democratic deliberation depends on citizens' capacity not merely to articulate fixed positions but to engage reflexively with alternative perspectives: a capability that requires comfort with provisional judgment and openness to revision. ‘Democratic deliberation... depends on participants' willingness to regard their own positions, interests, and beliefs as provisional and open to revision in the face of appropriate reasons’ (Knight and Johnson, 2011).

This political value of uncertainty emerges most clearly in the capacity for ‘representative thinking’: the ability to imagine and consider multiple perspectives on contested issues (Arendt, 1961). This form of plural thinking depends on maintaining a degree of uncertainty about one's own perspective-a willingness to consider that other viewpoints may contribute valuable insights to collective decision-making. Such uncertainty does not represent epistemological weakness but rather facilitates the expanded imagination necessary for robust political judgment.

Contemporary communicative environments systematically undermine this form of uncertainty by rewarding expressions of epistemic closure rather than openness to revision. Social media algorithms privilege content that expresses certainty and moral indignation, while penalizing nuance and provisional judgment. The resulting communicative landscape creates incentives against precisely the forms of uncertainty expression most vital to democratic deliberation.

Specialized LLM interfaces designed as transitional conversational spaces could foster communicative environments that reward rather than penalize expressions of provisional judgment. By designing interfaces that recognize and validate expressions of uncertainty as productive contributions to dialogue rather than weaknesses to be minimized, these systems could help reconstruct infrastructure necessary for robust political deliberations.

This approach echoes Forst’s emphasis on the extent to which democratic legitimacy depends not on achieving consensus but on ensuring that citizens can participate meaningfully in the processes through which collective decisions are justified (Forst, 2011). By designing conversational systems that facilitate this process of mutual justification while recognizing its inherently provisional character, we could potentially enhance the quality of political judgment in democratic contexts. The challenge lies not in eliminating disagreement but in creating environments that support productive engagement with difference-environments that foster ‘agonistic respect’ rather than antagonistic conflict (Connolly, 2005).

The cultivation of these capacities for political judgment is crucial in contexts of deep pluralism. Democratic politics requires that ‘others are not seen as enemies to be destroyed, but as adversaries whose ideas might be fought, even fiercely, but whose right to defend those ideas is not questioned’ (Mouffe, 2005). This agonistic conception of democracy depends on

communicative environments that enable productive engagement with difference-a condition that requires specific forms of uncertainty expression.

Transitional conversational spaces thus represent not merely a technical innovation but a potential reconstruction of the communicative infrastructure necessary for democratic renewal. By designing interfaces that privilege certain forms of uncertainty expression, we might create environments that support the development of political judgment in ways increasingly marginalized within digital ecosystems.

5. Designing transitional conversational spaces

5.1 From Alignment to Democratization

The dominant approach to LLM design focuses on ‘alignment’-ensuring that model outputs conform to human values and expectations. This approach typically involves fine-tuning models based on human feedback to reduce the likelihood of outputs that are harmful, offensive, or misleading. While this approach addresses specific forms of harm, it often fails to acknowledge values' contested and evolving nature in pluralistic democratic societies.

An alternative and complementary approach focuses not on aligning models with a given set of values but on designing interfaces that support the democratic iteration of values. This approach recognizes ‘the democratic paradox’-the fact that democratic societies are characterized by ongoing contestation over the meaning of their fundamental values (Mouffe, 2000). Rather than attempting to resolve this contestation through technical means, 'democratization' approaches aim to create interfaces that support productive engagement with value pluralism.

This shift from alignment to democratization has significant implications for how we conceptualize the design of this specific category of LLM interfaces. Rather than focusing primarily on filtering or modifying model outputs to conform to predefined values, democratization approaches seek to make visible the contingent and contestable nature of value judgments, thereby supporting ‘democratic iterations’- the ongoing reappropriation and reinterpretation of normative principles through democratic practice(Benhabib, 2004).

5.2 Supporting the Spectrum of Moral Articulation

The ability to move between back and forth different stages of moral articulation -from tentative exploratory thoughts to fully formed public positions- depends on specific environmental conditions. Traditional democratic discourse has relied on conversational contexts that support different forms of moral exchange, from intimate discussions among trusted friends to formal public debates. This varied ecosystem allows for to-and-fro movements between ‘imperfect rationalisations’ and more fully articulated moral positions (Williams, 1981).

LLMs’ deployment introduces distinctive challenges to this articulation ecology by potentially flattening the rich variety of conversational contexts that characterize human

moral discourse. This environmental flattening occurs in two principal ways. First, through premature formalization, where tentative expressions are resolved too quickly into definitive statements; and second, through context collapse: the erosion of the distinctive conversational settings that traditionally support different modes of moral articulation, from exploratory to declarative.

This challenge of supporting the full spectrum of moral articulation is not unique to LLMs. Even face-to-face deliberations in large settings inevitably compromise some of the conditions that make small-group conversations conducive to moral exploration.[13] The question, then, is how technological mediation might best support the necessary interplay between private reflection and public deliberation, acknowledging that effective public discourse depends significantly on participants' prior opportunities for tentative exploration of moral concerns.

The design of transitional conversational spaces must therefore prioritize features that explicitly support different modes of moral articulation, particularly those exploratory modes that are increasingly marginalized in contemporary digital environments. This requires a fundamental shift in design priorities away from optimisation for certainty and toward creating conditions for productive engagement with uncertainty.

5.3 Ensemble Interfaces and Democratic Contestation

A promising approach for implementing transitional conversational spaces is the development of 'ensemble interfaces'[14]-specialized LLM interfaces that present users with outputs from multiple models trained on different datasets or with different parameters. Unlike traditional ensemble methods in machine learning (which typically combine or 'average' the sub-model outputs), these interfaces would aim to make visible the contingent and contestable nature of the value-loaded choices underlying the differences between each sub-model.

Ensemble interfaces could support democratic contestation by revealing value assumptions embedded in different training regimes, supporting perspective-taking across different value frameworks, and facilitating comparison of different normative approaches. This approach acknowledges the inherently pluralistic nature of democratic societies and seeks to support productive engagement with this pluralism (rather than attempting to resolve it through technical means).

---

[13] 'A procedure in which people fail to internalize the perspective of one another qualifies as democratic only in the most mechanical of ways: without properly registering what one another is saying, it will be not an exchange of reasons but merely a count of votes.' (Goodin, 2000, p. 84).

[14] The 'ensemble' reference is there 'to flag their borrowing from parts of 'ensemble models' techniques. These techniques rely on the parallel running of one learning algorithm (or 'base learner') on different data subsets. These techniques' degree of rigour depends in large part on the way these data subsets are selected (and subsequent outcome differences resolved): when combined with 'bootstrap sampling' methodologies, these ensemble techniques can help reduce the risk of overfitting. For our purposes, such ensemble techniques could be just as helpful as those relying on multiple and slightly different learning algorithms (they may have different constraints imposed on the optimisation process). What matters is that the resolution process is taken out: rather than combining the results of each 'base learner' (whether through 'voting', 'averaging' or otherwise), emphasis would be placed on documenting the differences / factors that lead to each of the base learners' outcomes' (Delacroix 2022b).

The technical implementation of ensemble interfaces would involve training multiple models on different datasets or with different parameters. These datasets could be curated to reflect different value frameworks or could be selected to represent different stakeholder perspectives on a particular issue. Users would then be presented with outputs from all models rather than just a single 'aligned' or aggregated output.

This ensemble approach serves a specific function in creating transitional conversational spaces: it transforms what might otherwise be perceived as authoritative pronouncements into visible interpretations, each grounded in particular value frameworks and training methodologies. By making these differences visible, ensemble interfaces create opportunities for users to engage in the process of 'democratic iteration' described earlier.

5.4 Participatory Infrastructure for Collective Moral Development

Beyond ensemble interfaces, transitional conversational spaces require 'participatory infrastructure'-technical and social arrangements that enable communities to collectively shape the conversational environments in which they participate. This infrastructure is essential when addressing non-quantifiable forms of uncertainty, such as ethical uncertainty, which cannot be adequately addressed through pre-deployment design decisions.

The challenge of communicating ethical uncertainty exemplifies why designer-centric approaches are inherently limited. Unlike epistemic uncertainty (which can be quantified through confidence scores) or aleatory uncertainty, ethical uncertainty stems from the inherently dynamic and contextual nature of human values. There is no single correct way to express such uncertainty-its communication must be iteratively refined through collective use and feedback.

Consider how LLMs might communicate uncertainty in contexts like healthcare, education, or legal practice. Should the system express doubt through hedging language, present multiple perspectives, verbalising confidence, or explicitly acknowledge value tensions? No designer, however skilled, can determine in advance which approach will best serve the complex needs of diverse communities. Even more importantly, the appropriate expression of uncertainty will necessarily evolve as normative understandings develop over time within these practices.

This recognition necessitates participatory infrastructure with three essential characteristics:

First, it would support 'tinkerability': the capacity for user communities to experiment with and iteratively refine different approaches to uncertainty communication. Rather than treating uncertainty expression as a design problem to be solved before deployment, tinkerable systems would enable communities to continuously evolve how their systems express doubt, ambivalence, and normative complexity based on collective experience.

Second, it would incorporate mechanisms for collective governance over uncertainty communication. Individual personalization is insufficient when dealing with ethical uncertainty, as its appropriate expression emerges through collective deliberation rather than individual preference. Participatory infrastructure would enable communities to collectively

determine which forms of uncertainty should be highlighted in different contexts and how they should be expressed.

Third, it would create spaces for ongoing discourse about the values embedded in uncertainty communication itself. When an LLM expresses uncertainty about a morally charged topic, that expression inevitably reflects normative commitments about what deserves emphasis or caution. Participatory infrastructure would make these normative dimensions visible and subject to collective deliberation rather than hiding them behind designer decisions.

The case of non-quantifiable uncertainty illustrates why participatory infrastructure isn't merely desirable but necessary. There simply cannot be a pre-determined, optimal solution to communicating ethical uncertainty that designers could implement in advance. The appropriate expression of such uncertainty must emerge through ongoing collective engagement, reflection, and refinement. This is not a deficiency to be overcome but a fundamental characteristic of how humans navigate normative complexity.

Technical implementation of such participatory infrastructure would involve ‘interactive machine learning’ techniques that place communities ‘in the learning loop’ (Amershi et al., 2014). Instead of merely approving or disapproving system outputs, communities would actively shape how systems express uncertainty through iterative feedback and collective governance. This approach acknowledges that addressing non-quantifiable uncertainty is not a technical challenge to be solved once, but a continuous social process that requires infrastructure specifically designed to support collective moral development.

6. Democratic Renewal and Capability Enhancement

6.1 The Architecture of Moral Attention

Traditional approaches to AI governance, focused primarily on oversight and accountability mechanisms, fail to address how these systems reshape moral perception itself. The lens of ‘moral attention’ provides a more fertile conceptual framework for understanding what is at stake in the design of transitional conversational spaces. Moral attention refers to our capacity to notice and respond to ethically significant features of situations-a capacity that is shaped by the environments we inhabit.

The design of attention-conscious architectures demands looking beyond conventional metrics of efficiency toward what might be called ‘perceptual infrastructure’: the technological and institutional arrangements that condition how communities develop moral understanding.[15] Such infrastructure operates not simply through explicit rules or principles but through ‘skilled intentionality’: the embodied, context-sensitive capabilities that underlie moral perception (Rietveld et al., 2018).

[15] ‘Rietveld, Denys, and Van Westen elaborate how perceptual capabilities emerge through active engagement with relevant affordances in one's environment, highlighting that this process is not merely individual but fundamentally social in character.’ (Rietveld et al., 2018).

These attention-conscious architectures would support the complementary relationship between internal reflection and external deliberation. 'Internal-reflective deliberation can never literally replace external-collective ones,' but the two modes can mutually enhance each other when properly integrated (Goodin, 2000). Technological systems designed to facilitate this integration would recognize that democratic deliberation depends on creating conditions where participants can internalize multiple perspectives before engaging in collective decision-making.

Crucially, such architectures must acknowledge that technological mediation cannot substitute for embodied ethical encounters. The 'politics of presence'[16] highlights how the physical encounter with diverse others introduces dimensions of meaning that transcend simulated perspectives (Phillips, 1995). Rather than attempting to replace these encounters, attention-conscious architectures would aim to create transitional spaces that enhance our receptivity to subsequent face-to-face interactions-scaffolding rather than supplanting the embodied dimensions of ethical life.

6.2 Supporting Democratic Capabilities

The perspective of capability enhancement offers a productive framework for evaluating transitional conversational spaces. Rather than focusing primarily on the content they deliver or the values they encode, this approach considers how these specialized interfaces might expand substantive democratic freedoms: the real opportunities people have to participate meaningfully in shaping their collective lives.

Three democratic capabilities prove particularly relevant in this context. First, interpretive capability: the ability to recognize and make sense of ethically significant situations across different normative frameworks. Second, articulative capability: the ability to express moral concerns in ways that others can understand and engage with. Third, deliberative capability: the ability to participate effectively in collective processes of moral reasoning and decision-making. Each of these capabilities develops through specific social practices that technological environments can either support or undermine.

Transitional conversational spaces would aim to enhance these capabilities by creating environments specifically designed to scaffold their development. Unlike conventional AI interfaces, these specialized spaces would measure success through the extent to which they expand opportunities for meaningful participation in collective sense-making practices.

This capability-focused approach helps clarify the distinction between transitional conversational spaces and other LLM applications. While many applications legitimately prioritize accuracy, efficiency, or user satisfaction, transitional spaces would be explicitly designed to support the development of democratic capabilities. They would constitute a specific subset of LLM interfaces intended not to supplant human judgment but to scaffold its

[16] 'The politics of presence' emphasizes how physical encounters with diverse others introduces dimensions of meaning that cannot be fully captured through abstract representation, highlighting the embodied nature of democratic deliberation.' (Phillips, 1995).

development in contexts where existing digital environments systematically undermine the conditions necessary for democratic deliberation.

6.3 Institutional Forms for Moral Development

The design of transitional conversational spaces raises questions about institutional form-how these specialized interfaces might be governed and situated within broader democratic ecosystems. From that perspective, we might conceptualize moral attention as a kind of collective resource that requires particular institutional arrangements for its cultivation and protection (Ostrom, 1990).

This institutional perspective highlights three key design considerations. First, scale: the appropriate size and scope for communities engaged in collective sense-making practices. Second, boundaries: how to define who participates in governance processes and how participation is structured. Third, monitoring: how to assess whether these specialized interfaces are successfully supporting democratic capability building rather than undermining it.

Addressing these considerations requires developing 'attention institutions'-social arrangements specifically designed to support communities in maintaining and developing their practices of moral perception. Unlike conventional regulatory bodies focused primarily on content moderation or harm prevention, these institutions would aim to create the conditions necessary for ongoing moral iteration at both individual and collective levels.

Crucially, the cultivation of such institutions would also support the preservation of adequate diversity in ways of seeing and understanding (Rosanvallon, 2007), thereby counteracting the homogenizing tendencies of digital environments and creating normative innovation spaces.

6.4 Beyond Technical Participation: Toward Democratic Ownership

The implementation of transitional conversational spaces requires moving beyond 'abstraction traps'[17] in algorithmic accountability-the tendency to treat technical solutions as sufficient for addressing fundamentally socio-political challenges (Selbst et al., 2019). This means recognizing the limitations of procedural approaches to AI governance that attempt to translate the rich complexity of moral perception into formal mechanisms of oversight and control.

The limitations of procedural approaches stem from a misalignment between the nature of moral perception and the design paradigms that dominate contemporary AI development. While moral perception depends on maintaining productive uncertainty and attentiveness to emergent ethical concerns, technical frameworks typically privilege optimisation and convergence. This misalignment manifests clearly in implementations of 'society-in-the-loop'

[17] 'The "abstraction trap" occurs when technical discussions of algorithmic systems fail to account for how these systems are embedded in complex social and institutional contexts, leading to solutions that address symptoms rather than underlying structural issues.' (Selbst et al., 2019).

machine learning[18], where participation is operationalized through value elicitation mechanisms that fail to engage the dynamic, emergent nature of moral understanding (Rahwan, 2018).

6.5 From Method to Structure: Transforming Development Paradigms

Recent initiatives in participatory AI development have begun to address these limitations by developing methodologies that support more substantive forms of democratic engagement. The STELA methodology explicitly tackles normative questions of 'whose values and norms AI systems should be aligned with, and how these choices should be made' through community-centered deliberation on language model outputs (Bergman et al., 2024). Their findings suggest that deliberative approaches can provide 'rich contextual insights for AI alignment' while enabling 'an inclusive process that is robust to the needs of communities.'

However, methodological innovations alone cannot overcome structural impediments within the political economy of AI development. These impediments manifest in the growing tendency to replace actual human participants with simulated ones in AI development processes-a substitution that fundamentally misconstrues the nature of democratic participation (Agnew et al., 2024). This creates a misalignment between 'the localized engagement of community-based participatory methods, and the globalized operation of commercial AI systems' (Young et al., 2024).[19]

Addressing these structural impediments requires moving beyond methodological refinements toward transformative changes in development paradigms. The distinction between 'fitting participation' and 'forging relationships' articulates this necessity-the former adapts democratic practices to existing technological paradigms, while the latter transforms those paradigms through sustained democratic engagement (Cooper and Zafiroglu, 2024).

6.6 Alternative Development Models: From Theory to Practice

The Masakhane NLP project demonstrates the feasibility of alternative development models through a grassroots initiative for machine translation of African languages (Birhane et al., 2022). What distinguishes this project is not simply consultation with stakeholders but fundamental restructuring of development processes.[20] The project operated through democratic procedures where 'meeting agendas were public and democratically voted on' and participants could move 'fluidly between different roles' rather than being confined to predefined categories of contribution.

---

[18] 'Society-in-the-loop approaches attempt to incorporate collective values into algorithmic systems through formalized participation mechanisms, yet often reduce this participation to preference aggregation rather than substantive democratic deliberation.' (Rahwan, 2018).

[19] 'The tension between local participatory methods and global commercial AI systems creates significant challenges for meaningful democratic governance of these technologies, requiring new institutional forms that can bridge this gap.' (Young et al., 2024).

[20] 'The Masakhane NLP project's grassroots approach to machine translation of African languages demonstrates how participatory development can restructure power relations within technology development, creating genuine avenues for democratic control.' (Birhane et al., 2022).

This restructuring of development processes echoes the need for ‘a departure from the traditional ML textbook narrative’ toward approaches that focus equally on ‘process improvements and collective exploration’ (Mundt et al., 2025). Mundt and others’ analysis highlights how constraints on more collaborative AI systems stem not merely from social challenges but from ‘prevalent technical foundations’-the underlying assumptions and practices that shape how machine learning systems are conceptualized and implemented.[21]

These alternative development models suggest the possibility of ‘democratic capability building’: technological development processes that enhance communities' capacity for evolving forms of moral perception. Rather than pursuing technical closure through algorithmic optimization, attention-supporting infrastructures would aim to maintain key features of democratic iteration[22]: productive tension between universal aspirations and particular perspectives, preservation of genuine difference in moral perception, and capacity for innovation in ethical understanding (Benhabib, 2004).

6.7 The Politics of Transitional Spaces

The development of transitional conversational spaces ultimately involves a political choice - a decision about what kind of technological environments we wish to create and inhabit. This choice is not merely technical but concerns the conditions under which collective moral understanding develops. It involves ‘practices of freedom’-the ongoing work of examining and revising the sociopolitical arrangements that shape our possibilities for thought and action (Tully, 2002).

This political dimension highlights the limitations of purely procedural approaches to AI governance. Rather than treating democratic participation as an optional add-on to technical development, the perspective developed here recognizes the connection between democracy and the collective capacity for renewed moral attention. This recognition suggests the need for technological systems designed not merely to encode fixed moral principles but to support the ongoing development of spaces where nascent ethical concerns can find expression before entering the arena of formal political contestation. Such spaces would acknowledge the essentially dynamic nature of moral understanding and design technological systems to support this dynamism.

This political perspective helps clarify why transitional conversational spaces constitute just one valuable direction for LLM development rather than a universal prescription. Not all technological interfaces need to function as transitional spaces-many legitimate applications prioritize efficiency, accuracy, or user satisfaction for specific purposes. The argument developed here is not that all LLM interfaces should be designed as transitional spaces, but

[21] Mundt and colleagues argue for 'a departure from the traditional ML textbook narrative' toward approaches that focus equally on process improvements and collective exploration, recognizing how technical foundations shape participation possibilities.’ (Mundt et al., 2025).

[22] ‘Democratic iterations’ involve the ongoing reappropriation and reinterpretation of normative principles through democratic practice, allowing communities to both preserve and transform their shared values over time (Benhabib, 2004).

rather that this specific design direction warrants serious exploration given the systematic erosion of conversational environments in contemporary digital ecosystems.

7. Conclusion: Towards Specialized Transitional Conversational Interfaces

The emergence of LLMs as conversational partners represents a qualitative shift in how technological systems mediate the sense-making practices that underpin moral understanding. Unlike previous communication technologies that primarily transmitted fixed content, LLMs dynamically engage with and shape conversational flows, thereby reconfiguring the environmental conditions within which moral perceptions achieve articulation.

This mediational transformation poses particular challenges for democratic practice because moral understanding develops not through abstract reasoning but through iterative processes of dialogic exploration. Contemporary digital platforms systematically undermine these processes by privileging algorithmic patterns that reward definitiveness over uncertainty, closure over exploration. The result is a communicative infrastructure increasingly hostile to the tentative articulations through which diverse moral insights typically emerge.

Current approaches to LLM development fail to address this challenge because they focus exclusively on epistemic uncertainty: the extent to which these systems can convey their outputs with some measure of reliability. This focus overlooks hermeneutic uncertainty, which concerns the interpretive character of moral understanding itself. Democratic contexts require technological systems that can sustain hermeneutic uncertainty as a productive condition rather than treating all uncertainty as a technical deficit to be eliminated.

The distinction between these forms of uncertainty suggests design principles for 'federated attention architectures'. Rather than optimising for convergent solutions that homogenize interpretive frameworks, such architectures could maintain productive tension between diverse ways of knowing while supporting communities' capacity to refine their perceptual practices through conversation. The technical implementation involves ensemble interfaces that make visible the contingent nature of different interpretive approaches rather than concealing these differences beneath apparently authoritative outputs.

The same technologies that have contributed to this democratic deficit could, through intentional redesign, help address it. However, the development of such capability-oriented systems cannot proceed through conventional paradigms that concentrate design authority within technical communities. The analysis developed here demonstrates that addressing hermeneutic uncertainty requires participatory infrastructure enabling communities to collectively determine how uncertainty should be expressed and engaged within their specific contexts. This necessity stems from recognition that appropriate uncertainty expression must emerge through iterative collective engagement rather than predetermined design solutions.

Such participatory infrastructure operates through three interconnected dimensions that acknowledge the iterative, social character of moral understanding: 'tinkerability' (enabling communities to experiment with and refine different approaches to uncertainty expression

based on collective experience); collective governance mechanisms that transcend individual personalisation toward community-determined approaches to interpretive complexity; and ongoing discourse about the normative dimensions embedded within uncertainty communication itself, recognising that expressions of doubt inevitably reflect value commitments about what deserves emphasis or caution.

The implementation of these participatory dimensions, however, cannot occur within arrangements that concentrate design authority within technical communities. The capability-oriented approach developed here necessitates institutional transformations that move beyond procedural accountability mechanisms toward arrangements designed to support communities in cultivating their practices of moral perception. Such institutional innovations represent a departure from conventional approaches to AI governance, which typically address symptoms of technological mediation rather than engaging with its systematic effects on moral understanding. The erosion of 'uncertainty-friendly' conversational spaces leads to the rigidification of perceptual habits, transforming the effort inherent in moral understanding from dynamic practice to static ideology.

These institutional considerations reveal why the question of LLM design becomes inseparable from broader questions about democratic sustainability in increasingly mediated societies. The diversity and vitality of the sense-making practices upon which democratic institutions depend requires ongoing refinement of the pre-reflective dimensions of moral perception through conversational engagement. As LLMs increasingly mediate these conversations, the intentional development of transitional conversational interfaces represents one concrete approach to addressing this challenge—not through the technical resolution of moral disagreement, but through the enhancement of communities' capacity for productive engagement with interpretive uncertainty.

The exploration of specialized interfaces designed to support the iterative development of moral articulation thus represents one direction warranting investigation (alongside alternative approaches to LLM development that serve different institutional purposes and social needs). The stakes involved extend beyond immediate design choices to the technological environments we create for the conversations that sustain moral understanding. These choices will determine whether digital mediation enhances communities' capacities for productive engagement with interpretive uncertainty or systematically erodes the conversational conditions necessary for such engagement.

References

Agnew, W., Shah, R., Zhou, F., Scheuerman, M. K., & Wallach, H. (2024). The illusion of artificial inclusion. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1-15. https://doi.org/10.1145/3613904.3642703

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine, 35*(4), 105-120. https://doi.org/10.1609/aimag.v35i4.2513

Arendt, H. (1961). *Between past and future: Eight exercises in political thought*. Viking Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216-9221. https://doi.org/10.1073/pnas.1804840115

Benhabib, S. (1992). *Situating the self: Gender, community, and postmodernism in contemporary ethics*. Psychology Press.

Benhabib, S. (2004). *The Rights of Others: Aliens, Residents, and Citizens*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790799

Bergman, S., Marchal, N., Mellor, J., Mohamed, S., Gabriel, I., & Isaac, W. (2024). STELA: a community-centred approach to norm elicitation for AI alignment. *Scientific Reports, 14*(1), 6616. https://doi.org/10.1038/s41598-024-56648-4

Birhane, A., Rostamzadeh, N., Srinivasan, R., & Tesema, G. M. (2022). Power to the People? Opportunities and Challenges for Participatory AI. *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1-13. https://doi.org/10.1145/3551624.3555290

Bohman, J. (2004). Expanding Dialogue: The Internet, the Public Sphere and Prospects for Transnational Democracy. *The Sociological Review, 52*(s1), 131-155. https://doi.org/10.1111/j.1467-954X.2004.00477.x

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., & Brunskill, E. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313-7318. https://doi.org/10.1073/pnas.1618923114

Bruns, A. (2019). Filter bubble. *Internet Policy Review, 8*(4), 1-14. https://doi.org/10.14763/2019.4.1426

Connolly, W. E. (2005). *Pluralism*. Duke University Press. https://doi.org/10.1215/9780822387084

Cooper, N., & Zarifoglu, A. (2024). From fitting participation to forging relationships: The art of participatory ml. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1-18.

Delacroix, S. (2022b), Diachronic interpretability and machine learning systems, *Cross-disciplinary research in computational law,* (1)2, https://journalcrcl.org/crcl/article/view/9

Delacroix, S. (2022a), *Habitual Ethics?*, Bloomsbury (open access)

Delacroix, S. (2025), Moral Perception and Uncertainty Expression in LLM-Augmented Judicial Practice, https://papers.ssrn.com/abstract=4787044

Delacroix, S., Robinson, D., Bhatt, U, Domenicucci, J., Montgomery, J., Varoquaux, G., Ek, C.H., Fortuin,V., He, Y., Diethe, T., Campbell, N., El-Assady, M., Hauberg, S., Dusparic, I., Lawrence, N. (2025), 'Beyond Quantification: Navigating Uncertainty in Professional AI Systems', forthcoming in *RSS Data Science and Artificial Intelligence*.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729-745. https://doi.org/10.1080/1369118X.2018.1428656

Forst, R. (2011). *The right to justification: Elements of a constructivist theory of justice*. Columbia University Press.

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, (25/26), 56-80. https://doi.org/10.2307/466240

Goodin, R. E. (2000). Democratic deliberation within. *Philosophy & Public Affairs, 29*(1), 81-109. https://doi.org/10.1111/j.1088-4963.2000.00081.x

Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2018). *Avoiding the echo chamber about echo chambers*. Knight Foundation. https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf

Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Harvard University Press.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Indiana University Press.

Kadavath, S., Conerly, T., Askell, A., Henighan, T., & Drain, D. (2022). Language models (mostly) know what they know. arXiv preprint arXiv:2207.05221. https://doi.org/10.48550/arXiv.2207.05221

Knight, J., & Johnson, J. (2011). *The priority of democracy: Political consequences of pragmatism*. De Gruyter. https://doi.org/10.1515/9781400840335

Kompridis, N. (2011). *Critique and disclosure: Critical theory between past and future*. MIT Press.

Lugones, M. (1987). Playfulness, 'World'-Travelling, and Loving Perception. *Hypatia, 2*(2), 3-19. https://doi.org/10.1111/j.1527-2001.1987.tb01062.x

Medina, J. (2013). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199929023.001.0001

Mouffe, C. (2000). *The Democratic Paradox*. Verso.

Mouffe, C. (2005). *On the political*. Psychology Press. https://doi.org/10.4324/9780203870112

Mundt, M., Lang, U., Chen, P. Y., Khana, U., Santos-Rodriguez, R., Song, S., Bengio, S., & Smith, L. (2025). The Cake that is Intelligence and Who Gets to Bake it: An AI Analogy and its Implications for Participation. arXiv preprint arXiv:2502.03038. https://doi.org/10.48550/arXiv.2502.03038

Murdoch, I. (1970). *The Sovereignty of Good*. Routledge.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Oosterlaken, I. (2015). *Technology and Human Development*. Routledge.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807763

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Phillips, A. (1995). *The Politics of Presence*. Oxford University Press. https://doi.org/10.1093/0198294158.001.0001

Putnam, R. D. (2000). Bowling alone : the collapse and revival of American community. London, Simon & Schuster. https://psycnet.apa.org/doi/10.1145/358916.361990

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14. https://doi.org/10.1007/s10676-017-9430-8

Rawls, J. (2005). *Political Liberalism*. Columbia University Press.

Rietveld, E., Denys, D., & Van Westen, M. (2018). Ecological-Enactive Cognition as Engaging with a Field of Relevant Affordances: The Skilled Intentionality Framework. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.3

Rosanvallon, P. (2007). *The demands of liberty: civil society in France since the Revolution* (Vol. 154). Harvard University Press.

Schudson, M. (1997). Why conversation is not the soul of democracy. *Critical Studies in Media Communication, 14*(4), 297-309. https://doi.org/10.1080/15295039709367020

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59-68. https://doi.org/10.1145/3287560.3287598

Tully, J. (2002). Political philosophy as a critical activity. *Political Theory, 30*(4), 533-555.

van Domselaar, I. (2022). "Inquiry and Imagination in Adjudication: The Case of Digitalisation." Neth. J. Legal. Phil. 51: 187. https://doi.org/10.5553/NJLP/221307132022051002 00

Van Grunsven, Janna (2022). "Enactivism and the paradox of moral perception." Topoi 41.2: 287-298. https://doi.org/10.1007/s11245-021-09767-w

Walker, M. U. (2007). *Moral Understandings: A Feminist Study in Ethics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195315394.001.0001

Williams, B. (1981). *Moral Luck*. Cambridge University Press.

Winnicott, D. W. (1971). *Playing and reality*. Psychology Press.

Young, I. M. (2000). *Inclusion and Democracy*. Oxford University Press. https://doi.org/10.1093/0198297556.001.0001

Young, M., Ehsan, U., Singh, R., Tafesse, E., Gilman, M., Harrington, C., & Metcalf, J. (2024). Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday*. https://doi.org/10.5210/fm.v29i4.13642