

Advancing Deliberation with AI: Architecture, Evaluation, and Simulation

Maurice Flechtner

Centre for Democracy Aarau, University of Zurich
Computational Social Science Lab, ETH Zurich
maurice.flechtner@zda.uzh.ch

Abstract

My doctoral research revolves around the future of deliberative democracy. In particular I am contributing to the ongoing discussions about designing and assessing scalable deliberative processes and the role of AI in deliberation. Combining deliberative theory, computational social science, and machine learning, my project asks how Large Language Models and adjacent technologies can help extend the democratic and epistemic benefits of deliberation to larger publics without eroding the conditions that make those benefits possible. My cumulative dissertation comprises three interlocking strands: (i) a normative–architectural framework for up-scaling deliberation under real-world constraints, (ii) an automated pipeline for measuring deliberative outcomes via the Deliberative Reason Index (DRI), and (iii) an empirical benchmark of LLM agents as deliberators.

Background and Motivation

Contemporary democracies face rising polarisation, declining institutional trust, and an information environment in which misinformation often circulates faster than reasoned exchange (Bennett and Livingston 2018). Deliberative democratic theory offers a partial response: reasoned discussion among free and equal citizens has been shown to reduce affective polarisation, increase political knowledge, and improve the quality of collective judgement (Fishkin 2018; Niemeyer et al. 2024). Yet classical deliberative formats remain small, costly, and slow, limiting potential impact and even raising legitimacy concerns when they are coupled to consequential decisions (Lafont 2019). Recent work argues that AI, and Large Language Models (LLMs) in particular, could help close this gap (Landemore 2024). By automating aspects like summarization, facilitation, discursive stratification (Dryzek and Niemeyer 2008) or even participation itself (?), LLMs could potentially offload some of the most labour-intensive work around deliberative processes. However, the normative and empirical conditions under which such tools support rather than undermine deliberation are not well understood.

My cumulative dissertation addresses this gap by treating digital deliberation as a problem whose solutions must simultaneously satisfy normative commitments (equality, in-

clusion, plurality, authenticity, reflection), epistemic constraints (the integration of diverse perspectives), and engineering realities (cognitive load, coordination cost, motivational dilution at scale). The central research questions are:

1. How can AI-supported tools help scale deliberation without compromising its core normative principles?
2. How can deliberative outcomes be assessed in a scalable, automated, and theoretically grounded way?
3. Under what conditions, if any, can LLM agents participate in deliberative reasoning?

Work Accomplished

Project 1: Upscaling deliberation under constraint. In joint work with Francesco Veri, we have developed a normative and design-oriented framework for scaling deliberation, i.e. massively increasing participant numbers. Scaling is desirable for several reasons: it extends deliberation’s individual benefits to wider publics (Fishkin 2009; Gastil et al. 2023; Muradova and Arceneaux 2021), improves epistemic quality through perspective diversity (Landemore 2013) and it mitigates questions of legitimacy of lottocratic deliberative assemblies (Lafont 2019). We formalise five deliberative principles across three analytical levels (structural, discursive, cognitive-reflective) and identify four predictable design tensions that emerge at scale: quality versus size, trust versus anonymity, diversity versus depth, and consequentiality versus engagement. We then propose three principle-preserving mitigation strategies: parallelisation of small deliberative units, discursive stratification to route under-represented perspectives into participants’ deliberative experience, and motivation via traceability and gamification. The framework is grounded in deliberative theory and coordination-cost models of group deliberation and is intended to define the design space within which the AI-mediated interventions of Projects 2 and 3 operate. A manuscript is currently under review.

Project 2: Automated measurement of deliberative outcomes. Defining the goal of deliberation has long divided deliberative democrats. Against consensus-based conceptions criticized by agonistic theorists (Mouffe 1999) the Deliberative Reason Index (Niemeyer et al. 2024), captures the growth of intersubjective consistency, i.e. the shared understanding of how underlying reasons and values relate to pol-

icy preferences. The DRI does not impose a correct solution and is therefore uniquely qualified to measure collective reasoning quality without prior assumptions. DRI is, however, costly to apply, as it depends on manually authored Q-concourses of statements representative of the public discourse on the given issue. If AI is going to be deployed in deliberation, the field needs an evaluation infrastructure that can actually detect whether deliberation is happening, and it should be easily deployable. I am currently developing AUTOCONCOURSE, a retrieval-and-generation pipeline that assembles issue-scoped multilingual corpora, organises them via anchor definitions and multilingual sentence embeddings, and synthesises concise neutral statements with LLMs under human control. The pipeline has already been piloted on several occasions, including a live Swiss citizens' assembly on healthcare costs (100 participants, drawn by stratified sortition from ~22,000 contacts), and was validated qualitatively against a human-authored DRI survey on Swiss food policy, recovering roughly half of all considerations and most non-general policy options. However, it still needs some refinement to be easily applicable to any DRI statement generation task.

Project 3: Benchmarking LLMs as deliberators. Existing benchmarks of LLM collective reasoning use verifiable tasks (Du et al. 2023) and are routinely cited as evidence that LLMs can deliberate. I argue that political deliberation is different from tasks that LLMs encounter in the aforementioned benchmarks due to the ill-structured nature of political problems and that deliberative capacities in the political sense should therefore be evaluated separately. My benchmark evaluates three core aspects of deliberative quality (procedural quality, deliberative reasoning quality, perspective diversity) across 1,980 simulated five-agent deliberations across 12 citizen-assembly topics and 11 frontier LLM configurations. I find that LLM groups match human procedural quality but produce only small, topic-dependent gains in intersubjective consistency, concentrated on tractable rather than ethically contested questions. LLM groups operate at roughly one-third of human starting diversity and reverse the human convergence pattern: where human deliberation integrates diverse views, LLM deliberation accumulates similarity. Engineering diversity through persona prompting does not restore the human dynamic but inverts which component of deliberative reasoning is updated. Additionally, these findings can further inform the social-epistemological debate on the epistemic authority of LLMs (Milella and Cabitza 2026; Hauswald 2025).

Planned Work

Building on Project 1, I am developing a deliberation platform with CORDA Democracy according to the proposed principles, which will form the basis for empirical validation of our normative guidelines. For Project 2, I am moving AutoConcourse from qualitative validation toward quantitative benchmarking on additional case-studies, completing the pipeline's evaluation process. Project 3 informs further research around Human-AI interaction in deliberative contexts, where I am currently conceptualizing an experiment

to evaluate the impact of AI involvement on DRI results.

Contributions and AIES Fit

My work speaks directly to AIES concerns about AI in democratic institutions, the epistemology of AI reasoning, and the evaluation of AI systems whose claims to legitimacy outrun their demonstrated capacities. Its three contributions are (i) a multilevel normative framework for scaling deliberation that names tensions current deployment regimes obscure, (ii) an automated, theoretically grounded measurement infrastructure (AUTOCONCOURSE) that makes deliberative reasoning assessable at scale, and (iii) the first three-dimensional empirical benchmark of LLM agents in deliberation.

References

- Bennett, W. L.; and Livingston, S. 2018. The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions. *European Journal of Communication*, 33(2): 122–139.
- Dryzek, J. S.; and Niemeyer, S. 2008. Discursive Representation. *American Political Science Review*, 102(4): 481–493.
- Du, Y.; Li, S.; Torralba, A.; Tenenbaum, J. B.; and Mordatch, I. 2023. Improving Factuality and Reasoning in Language Models through Multiagent Debate. arXiv:2305.14325.
- Fishkin, J. S. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford New York: Oxford university press. ISBN 978-0-19-957210-6.
- Fishkin, J. S. 2018. *Democracy When the People Are Thinking: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press. ISBN 978-0-19-882029-1.
- Gastil, J.; Årsælsso, K. M.; Knobloch, K. R.; Brinker, D. L.; Jr, R. C. R.; Reedy, J.; and Burkhalter, S. 2023. Deliberative Panels as a Source of Public Knowledge: A Large-Sample Test of the Citizens' Initiative Review. *PLOS ONE*, 18(7): e0288188.
- Hauswald, R. 2025. Artificial Epistemic Authorities. *Social Epistemology*, 39(6): 716–725.
- Lafont, C. 2019. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford University Press Oxford, 1 edition. ISBN 978-0-19-884818-9 978-0-19-188274-6.
- Landemore, H. 2013. Deliberation, Cognitive Diversity, and Democratic Inclusiveness: An Epistemic Argument for the Random Selection of Representatives. *Synthese*, 190(7): 1209–1231.
- Landemore, H. 2024. Can Artificial Intelligence Bring Deliberation to the Masses? In *Conversations in Philosophy, Law, and Politics*. Oxford University Press. ISBN 978-0-19-886452-3.
- Milella, F.; and Cabitza, F. 2026. Perceiving AI as an Epistemic Authority or Algority: A User Study on the Human Attribution of Authority to AI. *Machine Learning and Knowledge Extraction*, 8(2): 36.
- Mouffe, C. 1999. Deliberative Democracy or Agonistic Pluralism? *Social Research*, 66(3): 745–758.
- Muradova, L.; and Arceneaux, K. 2021. Reflective Political Reasoning: Political Disagreement and Empathy. *European Journal of Political Research*, 61.
- Niemeyer, S.; Veri, F.; Dryzek, J. S.; and Bächtiger, A. 2024. How Deliberation Happens: Enabling Deliberative Reason. *American Political Science Review*, 118(1): 345–362.