An MIT Exploration of Generative AI • From Novel Chemicals to Opera

Generative AI for Pro-Democracy Platforms

Lily L. Tsai[1] Alex Pentland[2] Alia Braley[3] Nuole Chen[4]
José Ramón Enríquez[5] Anka Reuel[6]

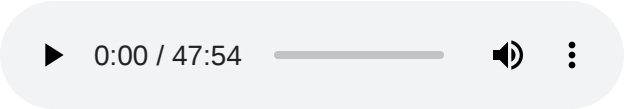
[1]Director and Founder, MIT Governance Lab MIT Political Science Massachusetts Institute of Technology,
[2]MIT Connection Science Research Initiative Media Arts and Sciences Massachusetts Institute of Technology,
[3]Department of Political Science University of California, Berkeley,
[4]MIT Political Science Massachusetts Institute of Technology,
[5]Stanford Graduate School of Business Stanford Institute for Human-Centered Artificial Intelligence Stanford University,
[6]Stanford Intelligent Systems Laboratory Stanford University

Published on: Mar 27, 2024

DOI: https://doi.org/10.21428/e4baedd9.5aaf489a

ABSTRACT

Online discourse faces challenges in facilitating substantive and productive political conversations. Recent technologies have explored the potential of generative AI to promote civil discourse, encourage the development of mutual understanding in a discussion, produce feedback that enables people to converge in their views, and provide usable citizen input on policy questions posed to the public by governments and civil society. In this paper, we present a framework to help policymakers, technologists, and the public assess potential opportunities and risks when incorporating generative AI into online platforms for discussion and deliberation in order to strengthen democratic practices and help democratic governments make more effective and responsive policy decisions.



Listen to this article

---

1. Generative AI for Pro-Democracy Platforms

In an era where opinion is a mouse click away, online discourse has become the defining crucible of contemporary ideas and ideologies. While social media platforms like Twitter, Facebook, and Reddit have shaped how we currently think of open discourse, these noisy, sprawling public squares are far from intentional, deliberative assemblies focused on solving problems. To paraphrase Taiwan's Minister of Digital Affairs, Audrey Tang, trying to have a political conversation on Facebook is like trying to have a political conversation in a nightclub [1]. In parallel to the rise, critique, and study of social media platforms and their effects on society, there has been a push toward building, studying, and deploying intentional technologies, including generative artificial intelligence (AI), to assemble individuals to share opinions on policy questions online and converge on recommendations. These 'deliberative platforms' use tools and technologies that surpass standard survey platforms by explicitly soliciting diverse perspectives on a given question, surfacing key comments for further examination by the participants and in some cases leading to collective decision-making.

These endeavors also extend beyond conventional methods of public deliberation. Historically, governments and communities have relied on approaches such as in-person town halls and open comment periods for regulatory decisions to enable citizens to share opinions and deliberate about policy issues. Outreach to citizens has been associated with higher trust in government and more citizen cooperation and engagement [2][3][4]. New online deliberative platforms promote themselves as technologies that can achieve these goals faster and at a larger scale, with less human bias and lower costs.

These new technologies have already changed the nature of what is possible in how discussion and deliberation on policy issues are achieved, and interest in experimenting with ways of integrating generative AI is now growing. Features of the first generation of these platforms range from open commenting, upvoting, and simple polling to organizing assemblies and participatory budgeting. Some platforms go even further, taking citizen opinions and using machine learning techniques to generate visualizations of the opinion space for users, such as areas of agreement, disagreement, and structural features intended to promote compromise or consensus.

These innovations raise important questions about how best to design these technologies and integrate generative AI so that they uphold individual rights and serve democratic objectives. How do we make sure that online platforms are themselves democratic as well as helpful to the overall democratic system? Given our values and objectives, the most promising directions involve developing these technologies to meet our current challenges such as polarization, mistrust, and democratic backsliding. In the following sections, we offer a framework for how to think about designing pro-democratic uses of generative AI, assess the opportunities and risks of current technologies in terms of this framework, and identify the most promising directions for integrating generative AI to support online discussion and deliberation.

2. A Framework for Pro-Democratic Generative AI and Deliberative Technologies

As citizens in a shared polity, we want online platforms to help us communicate our views to each other and arrive at wise collective decisions or recommendations. As with all technologies that affect the public, we want them to be transparent and accountable to the public. But what values and principles must we remain sensitive to if we want to design online platforms that take advantage of AI and other technologies to enable civic participation and discussion while also strengthening democratic societies? How do we make sure such platforms are pro-democratic and protect individual rights and political equality for all? At a minimum, our political traditions of liberal, republican, and democratic thought require us to design platforms that preserve user agency and autonomy; encourage mutual respect; promote fairness, equality, and inclusion; and augment our capacities for active citizenship rather than diminish them [5].[1]

Preserving Agency and Autonomy. Pro-democratic deliberative technologies and the use of generative AI must, as in other forms of democracy, preserve human agency and autonomy[6][7][8][9][10][11][12]. A person with autonomy and agency is capable of reflecting on their interests and goals and is free to make choices about taking action on these considered judgments without encumbrance or obstruction unless they are clearly doing harm to others. Technologies that distort or censor the information necessary for a person to understand how best to serve their interests and goals (for example, due to ‘hallucinations,’ biases in the training data or algorithmic bias), limit the actions they can take without the user’s knowledge, or deny a person the freedom to act on their interests and goals without compelling reasons to do so are anti-democratic.

For example, rather than picking an approach to visualizing different groups in a conversation on behalf of users, more pro-democratic platforms would instead inform users about different options, say, for how they would like to express themselves or how they would like to visualize the different groups in the conversation, and then enable them to choose from these options. Participants should also be able to adjust their preferences freely if they decide to align or compromise with a converging majority in order to reach consensus or decide to hold firm to their original preferences without manipulation or undue pressure or influence, for example, from elites, from extremist groups, or by bots or trolls.

Encouraging Mutual Respect. In the current environment of polarization and mistrust, pro-democratic platforms not only have to assemble individuals who may have diverse policy positions and enable them to converge on a recommendation, but they have to do so for people who may intensely dislike those who do not share the same political party or identity group. Encouraging people to respect fellow participants and their views becomes all the more important to keep top of mind when designing and evaluating AI-enabled technologies. For example, to preserve agency as well as encourage mutual respect, AI chatbots can check in with participants about language that is potentially offensive, prompt them to confirm whether they indeed find it offensive or not, and provide this feedback and/or alternative phrasings to authors upon their request [13].

Promoting Equality and Inclusiveness. Pro-democratic platforms should also uphold the principles of equality and inclusion [14][15]. In a democracy, people are meant to be political equals in the sense that they have the same right to participate and the same membership status in the polity. Each individual should thus have equal access to participating on a given platform. Once on the platform, every participant's voice should be heard, and each opinion should be counted in the same way. Elites or bots should not have the power to exclude certain users or groups. This is particularly important in situations where historically underrepresented groups or individuals who are directly or indirectly affected by policy outcomes have been systematically excluded from the decision-making processes that have led to the status quo. On the other hand, functionalities such as translation into different languages can improve equality and inclusion. Diverse members of the polity could have equal opportunity to express their views effectively and for others to understand them. Integrating AI could help by enabling dialogues between people in the same polity who speak different languages.

Augmenting Rather Than Substituting for Active Citizenship. As we think about the integration of generative AI into democratic platforms, we need to ensure that such technologies *augment* and *complement* human capabilities for taking action as citizens rather than *substituting* for them. Augmentation gives users more tools for obtaining and understanding the information they need to develop an informed position and for communicating their thoughts and positions effectively.[2] Platforms that provide translation into multiple languages, for example, augment and complement our capacities as citizens. Translation gives us another tool for communication so that we can express their views and read the views of others in our primary language. AI-assisted translation is a way of augmenting and complementing our existing capabilities rather than outsourcing or delegating to an AI something like voting that is core to our development and responsibilities as citizens.

Generative AI could also augment human capability by summarizing and sorting much greater amounts of information than an individual could on their own. High-quality information about the state of the world, the preferences of other individuals, and alternatives that participants might not have previously been aware of can promote reasoned, nuanced debate and help avoid oversimplification, as well as enhance interest in the discussion and the credibility of the platforms [16][17]. To the extent that such assistance helps users avoid paralysis from information overload, augmenting the ability of individuals to summarize large amounts of information could also support *agency*. At the same time, it is important to preserve *agency* by giving individuals choice over the breadth and depth of topics they access depending on their individual needs and preferences. Platforms must provide ways for users to understand why particular pieces of information or functionalities are available while others are not.

2.1. Enhancing Pro-Democratic Platforms through Deliberation

Pro-democracy discussion platforms, either implicitly or explicitly, aspire to ideals of 'deliberative democracy,' in which citizens come together with mutual respect as political equals to discuss political issues and make decisions about policies. Such deliberation entails "mutual communication that involves weighing and reflecting on preferences, values, and interests regarding matters of common concern" [18][19][20]. While deliberative democracy contrasts with aggregative democracy, which focuses on simply collecting and counting votes, these conceptualizations of democracy can also be complementary—deliberation may precede aggregative voting. The deliberative process involves people coming together to give and receive reasons for different policy positions with mutual respect, a focus on the common good, and a goal of generating collective decisions.

Participation in deliberation can yield benefits for individual participants, the quality of decisions, and the resilience of the system as a whole [21]. Participants ideally develop more tolerance, generosity, and empathy, more stable and coherent opinions, more justified preferences, and greater focus on the common good and public spiritedness, among many other benefits [21][22][23][24][25][26]. For instance, research has shown that participants in deliberative polling are more likely to learn, change their views, and participate [27]. Decisions resulting from deliberation may also be better for the collective good, as they are more likely to be based on a complete set of arguments for different options and take into account the vulnerable in society. Deliberation may also help identify areas where values converge and help sides reach mutually respected agreements even when consensus is not possible [28]. At the system level, deliberation helps to legitimize the resulting policies and decisions [29][30][31], including through the mutual justification of a decision through a collective process [32][33].

Designing pro-democratic technologies requires an attention to increasing the access to and availability of participation in deliberation—the extensiveness of deliberation—while trying not to weaken the substantive communication, interaction, and reflection—or intensiveness—that lead to deliberation's benefits for democracy. Both extensiveness and intensiveness matter, because research suggests that it is the *direct*

experience of participating in deliberation that has salutary effects rather than the observation or knowledge that decisions are made through deliberation [27][34][35]. Existing evidence also suggests that decisions from deliberation tend to be seen as legitimate by those that participated in the deliberation but not necessarily by those outside of the deliberation (see, for example, [36], [31], [37], [38], and [39]). Online deliberative platforms thus offer a compelling opportunity for increasing the extensiveness of deliberation by bringing more people into the deliberation process—but it must do so while maintaining the intensiveness of deliberation.

In Appendix 1, we provide comprehensive guidelines for assessing deliberative processes and outcomes, addressing the unique challenges posed by both online and offline deliberation. We emphasize the importance of evaluating these processes based on the principles of deliberative democracy included in our framework. The appendix describes what we consider a successful deliberation process, highlighting the need for respectful, autonomous participation and information-rich environments to foster active citizenship. It also underscores the significance of outcomes that are thorough, feasible, and supported by a plurality of participants, focusing on their practical implementation and the necessity of compromise. Additionally, we outline various metrics and methodologies for measuring deliberation and decision-making, ranging from nuanced speech analysis to approaches like quadratic voting, which enhance the richness of preference measures and provide a more detailed assessment of deliberation success.

3. Evaluating Current Deliberative Technologies

In what ways do current platforms promote the pro-democratic values and principles in our framework? In this and the following section, we use our framework to assess the ways in which the current generation of platforms might strengthen pro-democratic deliberation and discussion and to identify where generative AI might be most advantageous in this regard. A look at current technologies suggests that while many platforms have incorporated natural language processing (NLP) or sentiment analysis, only very few have begun integrating generative AI, although research and discussions in this area are accelerating. One reason for this is the lack of clarity on how these technologies can improve online deliberation processes.

Local, state, and federal governments around the world, as well as civil society organizations and citizen groups, are already making use of a range of deliberative platforms. Some are developed by nonprofits and others by for-profit organizations. Platforms also vary in whether or not they are open-source. The majority of applications we have seen are developed and used in the United States and France, as well as other countries in northern Europe, though there are also some examples of development and use in the Global South (See Appendix 2 for a list of platforms considered in this document).

A few of these platforms maintain the small group discussions that often characterize in-person deliberative meetings and use the platform to enable many of these small groups to interact via video or text and in parallel with each other. Most of these platforms, however, enable deliberation to happen across a larger period of time with a larger number of participants through asynchronous interactions. Some platforms solicit policy

proposals for collective deliberation from the crowd. Others enable deliberation on a specific policy question, sometimes proposed by a government body.

How pro-democratic are these current platforms in terms of each of the principles laid out in our framework? Many of them pay considerable attention to encouraging *mutual respect*, either through their content moderation functions or by limiting the number and nature of comments that participants are allowed to make. While web platforms do not have the best reputation for civility in discussion, civility is an essential element of the deliberative democracy ideal, and platforms have taken a variety of approaches to ensure or increase civility. A number of platforms do not allow users to respond to other users' comments directly, with the belief that this will mitigate trolling or personal attacks (Appendix 2 platforms 13, 20, 22, 24, and 27). Some do allow direct responses, but only by government officials (Appendix 2 platforms 2 and 10). Another structural feature some platforms use to generate civility is to constrain comments to 'pro' and 'con' columns rather than having a single comment thread (Appendix 2 platforms 6, 21, 24, and 27). Some platforms restrict the nature of comments even further by only soliciting questions directed to government officials, having required fields such as 'why the proposal is important' or by inviting only solution statements (Appendix 2 platforms 2, 13, and 20).

Of course, structural features do not guarantee that platform speech will remain civil, so most platforms have a moderation feature for administrators. But this approach relies on governments or organizations having enough people assigned as moderators, limiting the scaling potential of the platform. Some of the platform developers sell as a service part-time or around-the-clock human moderation (Appendix 2 platforms 5, 9, and 15).

A few platforms ask participants to exercise ***mutual respect*** as well as ***agency*** by actively choosing to make a commitment to respect others. These platforms ask users to consent to a pledge of civility before commenting (Appendix 2 platforms 6, 16, and 20). This is one area where some platforms have already begun to integrate NLP and sentiment analysis to flag toxic speech (Appendix 2 platforms 3, 25, and 27). In one particularly innovative approach, the Stanford Online Deliberation Platform bot listens for potentially toxic speech during small-group discussions and, if found, sends automated messages to all of the other group participants asking them to verify whether the AI is correct (Appendix 2 platform 25).

There is also increasing awareness about the need to design platforms that promote ***inclusiveness*** and ***equal access***. While web-based applications enable discussions at scale and for much larger swaths of a given population to participate in deliberation, at the same time, these technologies can exclude those without sufficient access or technical skills, which can be a particular issue for certain demographic groups and for countries in the global South. Some platforms attempt to increase inclusion by offering governments or civic leaders outreach services to representative samples of a given population or by specifying the number of participants needed to complete a deliberative process (Appendix 2 platforms 16 and 20).

One additional element that could improve inclusiveness and equal access can be gleaned from prior research on how to ensure that influencers cannot have the outsize influence that they have on regular social media, and prevent them from dominating the deliberation [40]. One effective way to do this is to constrain each individual to making only a few comments per day in a platform where identities must be registered and verified.

Some additional ways that current platforms address equality and inclusiveness include having built-in support for a finite number of languages (Appendix 2 platforms 4, 5, 16–18, and 25). Others are already using AI to provide translation between users (Appendix 2 platforms 3, 6, 9, and 27). In the only platform we found so far to have drawn heavily on large language models (LLMs), a subtype of generative AI that is used to generate text-based output, LLMs are used as part of a chatbot feature to help citizens write comments (Appendix 2 platform 3).

Another way that most current deliberative platforms work to treat participants equally and fairly is to both validate user identity while also allowing users to participate anonymously in the discussion.[3] This approach can reduce power differentials often present in in-person discussions due to a participant's identity—who they are, where they are from, what their socioeconomic status is, and other characteristics that might be deduced by personal information. This approach can also protect individuals from online and offline harassment, which can encourage more open and honest participation, free from the fear of personal judgment or reprisal. Validated but anonymous participation helps uphold ***autonomy*** and ***respect for persons***, and may be especially important for ***inclusiveness*** and participation from historically marginalized groups.

Finally, some of these platforms ***augment active citizenship*** by creating several new types of capacities for participants that can increase their understanding of what others believe, what they themselves believe, and where the group is most likely to reconcile their differences or remain divided in their views.

First, real-time summarization and visualization capabilities on these platforms can bolster the ability of participants to understand what the current status of the conversation is, what the major agreements and disagreements are, and what different subgroups of the population tend to think on a given issue. Beyond following new comments as they are being made, participants can look at real-time visualizations of the opinion space that show where current voting stands on a given issue (Appendix 2 platforms 13, 14, 18–20, 22, and 24). Some platforms go even further, using machine learning to generate graphical representations of themes and sentiments in the debate (Appendix 2 platforms 3, 4, and 22). In addition, a couple of platforms allow users to mouse over the visualization to view the most popular reasons given in a precise opinion area (Appendix 2 platforms 6 and 22).

For example, one platform, Consider.it, shows how the opinions of participants about a particular policy question are distributed on a spectrum from 'disagree' to 'agree' (Appendix 2 platform 6). Users decide how to filter data for visualizations based, for example, on demographic information like political party, or to weight

the data based on whether or not the person gave a reason for their opinion, whether their reason recognized tradeoffs (by contributing to the pro and con columns), or the amount of influence each participant has in the discussion. These visualizations increase capacity for active engagement as well as the ***agency*** to choose from a range of different visualizations. Greater understanding of others' opinions reduce polarization and increase commitment to democratic principles [41].

Second, tools and exercises such as decision trees and other visualizations provided by some platforms can help individuals clarify and reflect on their *own* views about a complex policy issue. Both OpenStad and Ethelo begin with an administrator adding in all of the possible policy configurations and tradeoffs in a given policy area (Appendix 2 platforms 16 and 21). The application then asks users a variety of smaller questions, which position them on a matrix or area of the opinion space, allowing them to adjust their preferences once they see likely outcomes or tradeoffs involved in their preferences. By the time they are in conversation with other users, they may have a more well thought-out and reasoned position. One platform, Assembl, uses AI to suggest actions to users such as providing a reason for their comment, giving an example, or stating their comment in more general terms (Appendix 2 platform 3).

Third, some platforms identify particular proposals or parts of the discussion as being the main points of consensus or controversy to help people avoid 'spinning their wheels' in a discussion that gets stalled. These functions help participants clarify and create mutual understanding of where conflict may be intractable, or alternatively, where the discussion is close to consensus (Appendix 2 platforms 6, 4, 20, and 22). Some platforms go a step further by highlighting an issue that would need to be addressed in order to gain consensus on controversial topics and allowing users to iterate on addressing those concerns (Appendix 2 platforms 6, 14, and 22).

Another related function is the flagging of similar proposals or comments to avoid splitting votes. Many platforms leave it up to users to decide what actions to take but offer them the opportunity to first search previous comments, tag comments with keywords, or propose to merge ideas (Appendix 2 platforms 2, 7, 13, 17, 20, and 22). Some platforms already use AI to suggest groupings of proposals by common themes or to suggest keywords in comments (Appendix 2 platforms 3, 11, and 18). These platforms tend to preserve human agency by requiring or at least allowing human confirmation of the AI-based categorization.

One of the most criticized aspects of online discussion in typical social media forums are the ways in which algorithms may be making politics more divisive. In response, some developers use their own algorithms to rank comments in a way that will facilitate civility or consensus, such as upranking comments with the most support (Appendix 2 platforms 2, 18, and 27), upranking comments with the most positive sentiment (Appendix 2 platform 20), or by using more complex algorithms that consider likelihood to generate support across difference (Appendix 2 platform 22), which is another place where AI-based approaches have started to be integrated.

In the latter cases, platform developers make their own decisions about how algorithms will rank comments and serve them to participants. Given that these algorithms are likely to be less transparent, these approaches to encouraging agreement or civility may fail to preserve human agency, the first principle in our framework. A number of platforms take a more pro-democratic approach, which allows users to choose how they want to sort comments, for example by most recent, most popular, or random order (Appendix 2 platforms 6–9, 11, 13, 14, and 24).

3.1. Example: Case of Polis

To illustrate some of these trends more concretely, we can take a closer look at one particular platform, Polis. Polis seeks to give citizens a dynamic overview of the entire spectrum of opinion around a discussion topic and has been seen as a highly effective direct and deliberative democracy social media tool [42]. It allows the government to pose policy questions to the public and then uses statistical summarization to provide graphical feedback on what the population as a whole believes or desires. The system claims to be effective at achieving popular consensus around contentious issues over a period of two or three weeks with anywhere from 100 to tens of thousands of participants or more.

Polis has been used to generate consensus on climate issues in Austria (2022), in Uruguay on a national referendum (2020–2021), in New Zealand to facilitate the development of government policy (2016–2019), in the Philippines to generate municipal policy (2020–present), in the US to counteract polarization in a Kentucky town (2018), in the UK as a part of a government polling effort (2020), and in Germany to develop a political party's platform (2018) [43]. In addition, it has been used for Twitter's Community Notes and by Anthropic to draft a publicly sourced constitution for an AI system [44][45]. Taiwan's deployment of Polis (divya-vtaiwan) is widely believed to be the most effective example of achieving popular consensus around contentious issues [42]. While case studies featuring the achievements of Polis and other deliberative and direct democracy platforms are promising, it is unclear to what extent failures of these platforms in practice have simply not been documented to the same extent.

The way Polis operates is that a topic is put up for debate. Anyone with an account can post comments on the topic, and can also upvote or downvote other people's comments. Unusually for online media, users cannot reply to other users' comments, making it difficult to engage in trolling. The upvote/downvote mechanism creates a citation network, similar to citation networks used in scientific papers, patent applications, and legal decisions, in which the upvotes and downvotes are analogous to citations. Polis does not use NLP or the identification of topics raised within the content of comments. Instead, this network of 'citations' drives citizen interaction on the platform.

This citation mechanism enforces an important constraint that is likely critical to the success of the system. The process of surveying comments made by others for upvoting and downvoting forces people to learn about

others' opinions, which has been shown to reliably promote 'wisdom of the crowd' effects and better decision-making [46].

The Polis system also uses the upvotes and downvotes to generate a citation map of all the participants in the debate, clustering together people who have voted similarly. Although there may be hundreds or thousands of separate comments, like-minded comments cluster together in this map, showing where there are divides and where there is consensus. According to the theory of how Polis works, participants then naturally try to draft comments that will win votes from both sides of a divide, gradually eliminating the gaps.

The Polis visualization of the comments, as shaped by citations, seems to be very helpful in promoting convergence of opinion, and is much like the visualizations that have proven very effective in domains such as finance [47]. In this regard, Taiwanese Minister of Digital Affairs, Audrey Tang, declared, "If you show people the face of the crowd, and if you take away the reply button, then people stop wasting time on the divisive statements" [48].

MIT research has shown that there is reason to believe that the Polis-style approach could have a very significant impact on decreasing polarization. We use this type of approach to achieve significant increases in democratic attitudes among partisans in America [41], and recent related MIT papers show convergence of opinion in financial decisions by providing users with a visualization of the range of opinion and action [46] [47][49]. MIT research also shows that outreach to citizens is associated with higher trust in government and higher levels of citizen cooperation and engagement [3][50].

3.2. Assessing the Risks and Benefits of Generative AI

Much of the interest in generative AI for deliberative platforms involves helping citizens to discuss more respectfully and identify points of agreement, producing feedback that may help people to converge in their views and summarizing citizen input on policy questions for use by government officials. Researchers at DeepMind, for example, have fine-tuned an LLM to generate statements about policy issues that maximize the expected agreement for a group of people with diverse opinions. They find that study participants in the UK prefer the consensus statements generated by their best model to statements written by people more than 65 percent of the time, suggesting the potential for LLMs to help diverse groups of people find agreement [51].

More modestly, generative AI can help identify likely areas of consensus before deliberation and argumentation begins, which could expedite the decision-making process and reduce unnecessary conflict (although to preserve autonomy from undue influence, it may be advisable to communicate these opportunities for consensus after participants have had a chance to develop their initial views). Generative AI has the potential to enhance the thoroughness of deliberations, help participants make proposals that are implementable and focus on achievable goals, and streamline the process for a group that wishes to do so to come to an agreement.

However, these sorts of digital innovations raise questions about how to mitigate potential risks and harms. For instance, AI-generated consensus statements may create agreement because the statements are written in a more friendly or more authoritative tone and not because people really share the view expressed. AI moderation might discourage diverse preferences, especially if the AI is trained on biased data sets. Using generative AI to increase the speed with which the group finds areas of agreement or possible compromise might reduce creative, genuine, and engaging discussion and reflection, as the focus shifts to efficiency and agreement rather than the exploration of diverse ideas or the practice of constructive argumentation. Finally, an over-reliance on AI could lead to an erosion of trust in the platform and deliberative process, as participants may feel that their contributions are undervalued or overshadowed by algorithmic decision-making.

When participants are not aware of these effects or the integration of AI into the platform more broadly, it also calls into question whether the principles of ***agency*** and ***respect*** are really being upheld. Manipulation by bad actors who might exploit AI systems to skew discussions or disseminate misinformation is a legitimate concern [52][53]. AI-generated deepfake images and videos have become an insidious problem on social media and news platforms [54][55][56].[4] Bots with social media accounts disseminate and spread misinformation, with one recent incident involving a GPT-3 bot posting on Reddit.

With the increasing use and advancement of generative AI, the generation of misinformation may become more effective and automated; experimental setups such as CounterCloud have shown that the generation and dissemination of misinformation may become more streamlined and effective in the near future [57]. This increased sophistication may have significant negative effects on online deliberative platforms. It is, therefore, critical to be aware of and mitigate the likelihood of these systems being exploited by bad actors and to avoid subjecting people to persuasion unknowingly.

There is also the risk of over-censorship or differential censorship to ***agency*** and ***equality***, where AI-based features might disproportionately silence certain viewpoints, either due to inherent model biases due to algorithmic design decisions, e.g., which data it's been trained on, or manipulation by external entities. Moreover, AI systems may struggle technically to adapt to emerging or rapidly growing topics, especially if they are out of their training data distribution, limiting their effectiveness in dynamic or evolving policy conversations and contexts. Protecting and preserving minority interests and views remains important.

Concerns about the lack of transparency in algorithms and the potential for infringements on freedom of speech cannot be overlooked. Processes that develop trust in the moral character of other participants and the hosts of the deliberation are thus essential. When citizens lose faith in the intentions and responsiveness of authorities, they may stop cooperating with each other [3][50][58][59].

Finally, participants may become too dependent on AI, leading to a lack of critical engagement with the issues under discussion and with the views and arguments expressed by other individuals. Generative AI should help us look for information as we seek to educate ourselves about the issues rather than doing all the work for us.

To uphold this fourth principle in our framework, we need to monitor closely the extent to which AI helps increase the *extensiveness* of participation, making it easier for more people to participate in collective conversations and democratic deliberation about policy decisions at the expense of the *intensiveness* of their participation in terms of the time, effort, and attention devoted to becoming informed, reflective, and empathetic citizens.

4. Recommendations for Future Directions

Our framework seeks to prompt the questions we should ask ourselves as we integrate generative AI in online platforms for discussion and deliberation. Generative AI should assist citizens without reducing their agency. It should treat citizens as political equals and enable citizens to treat each other with mutual respect. It should both protect citizens from harm by biased algorithms or bad actors while also engendering the trust essential for participation in democratic deliberation. And AIs should not perform our responsibilities as citizens on our behalf or serve as our representatives in policy deliberation or policymaking processes.

In our own research, we are now beginning experimental evaluation of both the original Polis system and generative AI-supported versions of the system [60]. We expect that our investigation will provide insight into how to build an effective digital democracy that is consistent with the principles in our framework, while also clarifying the decisions we need to make about balancing these principles and protecting against the aspects of generative AI that are the most dangerous for democratic deliberation.

Specifically, we see the potential to improve online deliberative democracy initiatives using generative AI to support the *intensiveness* of user engagement and deliberation on the platform as well as increase the *extensiveness* of deliberation through the scaling of such platforms while upholding the democratic commitments to preserving human agency, mutual respect, equality, and inclusiveness and augmenting for active citizenship laid out in our framework for pro-democracy platforms. The former directly helps users to improve the depth and thoughtfulness of their participation on the deliberative democracy platform, e.g., by using LLMs to provide comment-writing support or reliable information. The latter supports the scaling of such platforms to include and integrate participation from more people, which has been a challenge for traditional deliberative democracy platforms. Building on the basic structure of current deliberative democracy platforms (as described above), we believe that the following four strands of generative-AI–based expansions are particularly promising and should be tested in corresponding experiments.

4.1. Intensiveness of User Engagement and Deliberation

Reflection and Writing Assistance. Generative AI tools may also help promote convergence of citizen opinion by encouraging communication with mutual respect and inclusiveness. Polis-like platforms try to get participants to reflect more carefully and engage more deeply when they write comments, thus improving the *intensiveness* of deliberation. Features of such comments are that they are non-toxic, well-reasoned, and

oriented toward the public good and public interest. LLMs could potentially support participants in the commenting stage to write comments that fulfill these criteria.

For instance, with the permission of the participants, LLMs could be used to analyze participants' comments and ask probing questions or provide substantive feedback in order to nudge participants to increase nuance and focus on well-reasoned comments. Similarly, an LLM could be used to nudge participants to consider public interest and public good by asking questions like 'how would this affect the budget?' or 'how will this change the role of police in the community?' It is promising that such AI tools are already deployed on many social network platforms, where users readily accept these terms for using and participating on the platform [61].

4.2. Extensiveness and Scaling of Deliberation

Summarization and Support for Decision-Making. In written deliberations, LLMs could be used to summarize the current discussion, distill themes, highlight points of controversy, and provide road maps toward potential consensus. For instance, they could be used to inform people about how they might generate policy proposals with more widespread support or acceptance, increasing the *extensiveness* of the deliberative process, based on the analysis of the ongoing discussion and the evaluation of the most-supported comments.

Visualization. Another avenue for improvement is the visualization of community opinion. For instance, in our recent paper, the analysis of science, patents, and law citation networks (which are very similar to the citation network used in Polis) allows for the visualization of the evolution of the networks, making it easy to predict convergence of the community around certain views [62]. We expect the same sort of visualization to be effective in helping citizens understand the evolution of opinion on a Polis-like platform. Generative AI approaches could be used in a first step to summarize and classify opinions; this information could subsequently be used to provide more flexible, automated visualizations that users can adjust based on their preferences. This is likely to be a place in which AI should augment rather than substitute for moderation because, in their current form, generative-AI–based visualizations suffer from bias and reliability issues.

4.3. The Need for Rigorous Research

There are further potential downfalls of using generative AI in deliberative online processes. For instance, using LLMs to support participants' writing processes may have negative implications, such as a reduction in content diversity and bias, such as when trying to detect toxic content or suggest alternative writing. Depending on the data these models have been trained on, they may exhibit political biases themselves, as shown in a recent study [63]. There may also be limits with respect to using LLMs for low-resource languages, where detection and suggestion capabilities might be significantly worse than in English. Hence, generative AI tools for use as deliberation aids must be very carefully crafted and tested, and their performance continuously audited; a specific focus when evaluating these models should be put on vulnerable or minority groups who have traditionally been disadvantaged in democratic and deliberation processes. Our framework for designing

pro-democratic AI can help guide the research that must accompany experimentation with new integrations of generative AI into these platforms.

Further research will also be needed to assess the tradeoffs between intensiveness and extensiveness. While we want deliberation to include more people and to be representative, we also want their participation and interactions during deliberation to be meaningful and trusted and to stem from reflection about the issues and understanding of each other. The choice of one design might make deliberation more extensive at the cost of rendering it less intensive. Research can help us identify these trade-offs, weigh the costs and benefits, and decide on how much we need of each.

4.4. Concluding Remarks

Meaningful participation in respectful deliberation can be transformative. It can rebuild the trust we have in democracy and in each other. Online platforms and generative AI give us extraordinary new opportunities to participate in discussions and policy deliberations with each other at scale.

The question we ask is not what can we do with these technologies, but what should we do? AI will make its way into our political processes whether we like it or not. This paper suggests the guardrails for how we make this grand experiment strengthen rather than further degrade our democratic systems. How should we use these technologies to promote the collective well-being of all citizens, our individual rights, our status as political equals, and our dignity, agency, and development as human beings with responsibility for ourselves and each other? People have come to find too much of what is out there now as illegitimate and toxic. So far we have not held social media technologies to the principles highlighted in our framework. But it will be essential to do so as generative AI further amplifies the ability of online platforms to fortify—or sabotage—our democratic values and societies.

As Alexander Hamilton urged in *The Federalist Papers*, we must seek to establish democratic processes and good governance through reflection and choice rather than by accident and force. We know what we have doesn't work. The good news is that we have new capabilities to see what we can do differently and how we can do better.

Acknowledgements

Development of this white paper was supported by generous gifts from MIT and from Project Liberty.

Appendix 1. Evaluating Deliberation

In this section, we aim to provide actionable guidelines for researchers and practitioners on how to evaluate deliberative processes and their outcomes. Assessing deliberative processes comprehensively—both online and offline—poses considerable challenges. Since deliberative democracy is intrinsically different from other forms of democracy, we ought to evaluate whether and to what extent its foundational principles are being

realized using different criteria. For instance, the ‘one man, one vote’ principle set at the core of participatory democracy might acquire nuance within a deliberation arena as participatory democracy also aims to incorporate mutual respect, equality, and inclusiveness of opinions as highly-esteemed conditions, as previously described in our framework.

The implementation, assessment, and monitoring of deliberative democracy processes poses several challenges. While these challenges are certainly present for all forms of deliberation, they vary in their nature depending on whether deliberation occurs online—anonymized in some instances—or offline. In general, the main challenge revolves around balancing inclusiveness and efficacy. The former condition is presented as one of the core principles in our framework and the latter as an implication of the ‘augmenting citizenship’ principle also presented before . These two critical dimensions, each laden with its complexities, come to the forefront when evaluating such democratic procedures: the deliberation process itself and the resulting outcomes.

Metrics of Success

Assessing deliberation cannot be reduced to a simple binary measure of success. This holds true for both the deliberative process itself and potential resulting consensus, which would inform collective decision-making procedures. Instead, the effectiveness of deliberation can be gauged by dissecting and evaluating various pertinent aspects of the process and the attributes of the attained outcome. With respect to the processes, these factors encompass elements such as preserving agency and autonomy, as well as creating information-rich environments to foster active citizenship. In terms of outcomes, they involve considerations like thoroughness, feasibility, and compromise.

Deliberation Processes

Successful deliberation processes ensure that individuals demonstrate respect for others’ opinions, act voluntarily (i.e., enjoy autonomy), and have access to the resources needed to make informed decisions so that they engage in active citizenship. Firstly, successful deliberation takes place in the public sphere, in environments where participants stand on equal footing and consider the common good [64][65][66].[5] This ensures that every participant’s voice is heard and that each opinion carries equal weight. This is particularly important in situations where historically underrepresented groups or individuals who are directly (or indirectly) affected by policy outcomes have been systematically excluded from the decision-making processes that have led to the current status quo.

Secondly, effective deliberation thrives in contexts where individuals enjoy autonomy. The influence of elites or group capture, in any form, can undermine the process. Autonomy is therefore essential for individuals to learn about different alternatives that they might not be aware of initially. Autonomy is also relevant for individuals to be able to freely and convincingly adjust their preferences if they decide to align—or compromise—with a converging majority.

Thirdly, information-rich but non-saturated environments are crucial for individuals to arrive at optimal outcomes and to safeguard their autonomy. High-quality information promotes informed and reasoned debate over simplification and misinformation. Given limited attention and energy, individuals in information-polluted environments are likely to arrive at false consensus or be compelled to make decisions based on suboptimal alternatives. Therefore, putting emphasis on reasons (also referred as 'justifications') [64] or emotions that convey identity features in contrast to asserting opinions without justification can enhance the thoroughness and engagement of deliberative outcomes. Finally, sharing high-quality information about the state of the world or the preferences of other individuals not only increases knowledge about other participants and the specific issues under discussion but also enhances interest in and credibility of the entire deliberation process [16][17].

Deliberation Outcomes

When assessing the success of deliberation outcomes, it is essential to abstain from considering the content of the outcome itself. Instead, in addition to evaluating the conditions under which consensus could have been achieved and how participants' decisions are made, the focus should be on the thoroughness and feasibility of the outcome as well as the support for it.

First, consensus is more valuable when the alternatives under consideration are well-specified and the tasks that result from the outcomes are well-defined. Unlike other democratic processes, except perhaps for participatory democracy, deliberative democracy's outcomes must be clear to enhance tracking and accountability among participants.

Second, the outcomes must be realistic and feasible. While consensus is more likely to occur for less concrete and more idealistic alternatives, deliberation is more useful when considering viable alternatives. Actionable tasks are therefore necessary to translate individual stated preferences into actual collective decisions.

Finally, outcomes resulting from consensus must enjoy minimal support. This means that when individuals have agreed upon an alternative, they must compromise and not actively oppose it in the future. Ongoing commitment and accountability for the outcome, based on the well-specified tasks, is more likely to be upheld when individuals who initially did not favor the alternative offer (minimal) support or compromise.

Finally, in the next two subsections, we present some guidelines for measuring deliberation and consensus. The objective is to lay out multiple approaches that have been used previously in the literature as well as innovative approaches to do so.

Measuring Deliberation

Nuanced measures of speech can be used to characterize the deliberation process as a whole. An outstanding example of this type of measure is the Discourse Quality Index [66], which includes categories reflecting how well a discourse aligns with some of the relevant theoretical principles of democratic deliberation outlined earlier [64]. Other examples include quality of deliberation measures [67][68][69], which encompass relevant

dimensions such as equal participation, respect for opinions, adoption of societal perspectives, and reasoned justifications. While typical applications of these measures include parliamentary debates, these measures can all be adjusted for online and offline deliberation among citizens as well.

Apart from fully-encompassing measures of deliberation, there are some alternative frameworks evaluating deliberative processes. Such frameworks are based on particular desired characteristics that the process entails or the outcomes provide. The most prominent of these principles are inclusivity and effectiveness.

A first crucial aspect in assessing the success of deliberation processes is the representation of a diverse range of voices, with a specific emphasis on the inclusion of minorities. This aspect of deliberation stresses the importance of ensuring that all segments of society, especially those that are often marginalized or underrepresented, have a voice in the decision-making process [14][19][70].[6] Moreover, measures such as the 'deliberative uptake' measure have been developed to gauge the inclusiveness of these processes [71]. Apart from increasing inclusiveness, ensuring that a multitude of perspectives, especially from minority groups, are heard and considered, deliberative democracy processes enhance the legitimacy and acceptability of the outcomes.

Finally, the effectiveness of deliberation is a critical measure of its success. This dimension is characterized by the acceptance of the deliberative outcomes and the ability to reach reasonable and consistent conclusions. Effective deliberation is not just about the process of discussion but also about achieving outcomes that are seen as legitimate and well-founded by all participants. It involves reaching decisions that reflect a collective understanding and acceptance, even if there is not complete agreement. Most of the work measuring effectiveness of deliberative processes has been centered around measuring the extent to which it manages to integrate diverse viewpoints into a coherent, transparent, credible, and mutually acceptable resolution (see, for example, [72][73]).

Measuring Consensus

Consensus has received vast attention both in the literature and in on-the-field deliberation processes. While compromise among participants is generally advantageous for decision-making processes, particularly if the objective is producing a policy recommendation, we believe it is not a sine qua non condition to determine the success of deliberation processes. In fact, achieving consensus is only desirable if both (1) it broadly captures the will of the group and (2) it respects the preferences of the minorities and the participants directly affected by the group determination. Therefore, it is also important to consider that consensus might come at the expense of diversity, one of the principles in our framework.

On the extensive margin, consensus is achieved when all members of a group agree upon an alternative. However, consensus is subject to multiple nuances, including the diversity of the original preferences and the depth of the consensus. This introduces a range of possibilities for evaluating consensus, which are associated

with the previously discussed elements of success. Below, we present a few metrics that we believe are relevant for assessment, especially in the context of potential interventions facilitated by generative AI.

When assessing whether consensus has been achieved, two types of options emerge, both based on stated preferences. These preferences can be derived from comments, reactions, or even votes. The first set of metrics estimates the extent of preference discrepancy, with no discrepancy indicating consensus. The most basic statistic for this category of metrics is variance, where a value of 0 represents consensus. Another option, especially when there are other distinct and relevant choices, is the Herfindahl index, where a value of 1 characterizes consensus. Lastly, a simple ratio between the relevant options available to participants (e.g., likes and dislikes) as the share of the total available options can be used to gauge dissent or consensus.

The second set of metrics pertains to the number of relevant dimensions within a conversation or a voting procedure. Many of these measures have been developed to reduce the dimensionality of written texts but can also be applied to short snippets of text derived from transcripts of online conversations or voting. These metrics may encompass the number of topics or relevant factors and clusters, as determined by methods like principal component analysis, *t*-distributed stochastic neighbor embeddings, *k*-means and hierarchical clustering, unsupervised versions of Latent Dirichlet Allocation models, Structural Topic Models, or non-negative Matrix Factorization. Additionally, sequence modeling techniques like hidden Markov models and recurrent neural networks can be considered.

Even more importantly, consensus is seldom reached without nuances. Therefore, it is crucial to gauge the intensity of participants' preferences in deliberative processes. By enhancing the richness of preference measures, we can move beyond evaluating mere consensus and better assess the success of deliberation, as detailed earlier. Ultimately, these measures help distinguish between those participants willing to compromise as well as the collective minimal level of understanding and acceptance.

On the intensive margin, consensus can be approximated by examining the intensity of preferences. This can be achieved through various approaches like quadratic voting, graded preference voting, or cumulative voting [74][75][76][77].[7] Votes or comments can serve as input alike. The ability to consider a more nuanced relative ranking of options can help (1) determine Pareto-improving alternatives and (2) identify areas where it might be beneficial to enhance task monitoring to ensure completion of the consensual outcome.

Appendix 2. Direct and Deliberative Democracy Platforms

This list represents the deliberative platforms considered in this document. While we cannot claim to have created an exhaustive list, we have tried to be as comprehensive and inclusive as possible.

Table 1

Ref. #	Platform Name	Developers

1	Adhocracy+	Liquid Democracy
2	Ask Parliament	Democracy Developers
3	Assembl	Bluenove
4	Citizen Lab	Citizen Lab
5	Civocracy	Civocracy
6	Consider.it	Consider.it
7	Consul Democracy	Consul Democracy Foundation
8	ConsultVox	Publilegal
9	CoUrbanize	CoUrbanize
10	Crowd Law Making	Democracia en Red
11	Decidim	Decidim Free Software Association
12	Decision21	Participation21
13	Dialogue	Delib
14	Discuto	Impressum
15	EngagementHQ	Granicus
16	Ethelo	Ethelo
17	Fluicity	Fluicity
18	LiquidFeedback	LiquidFeedback
19	Loomio	Loomio
20	Make.org	Make.org
21	OpenStad	Municipality of Amsterdam
22	Pol.is	The Computational Democracy Project
23	Public Consultation	Democracia en Red

24	Purpoz	Cap Collectif
25	Stanford Online Deliberation Platform	Deliberative Democracy and Crowdsourced Democracy Teams at Stanford
26	Wikum	MIT Researchers
27	Your Priorities	Citizens Foundation

Footnotes

1. In this article, we use the term ‘democracy’ to include what political philosophers refer to as electoral (i.e., having free and fair elections) and liberal (i.e., respect for human rights) elements. We refrain from using the term ‘liberal’ democracy so as not to create confusion with more politicized uses of the word ‘liberal.’ For more information on different elements of democracy, see [5]. ↩

2. Augmentation is related to but not the same as agency. While augmentation is about increasing our options for accessing and digesting information, preserving agency is about preserving our ability to make choices about how we receive our information, how we want to visualize it, and our policy position without coercion, undue influence, or manipulation. ↩

3. Most platforms seem to have the option to allow administrators to choose the level information users are required to share as well as their level of anonymity within the discussion forums. ↩

4. Some recent examples include an AI-generated image of a fake explosion at the Pentagon [54], AI-generated videos of the Ukrainian and Russian presidents declaring the war was over [55], and AI-generated images of the war in Gaza [56]. ↩

5. Participants should consider some sort of common good, even when they have limited information about their own best interest (see [64], [65], and [66]). ↩

6. Jane Mansbridge's work ([14], [19], and [70]) in this area underscores the significance of inclusivity in deliberative democracy. ↩

7. For quadratic voting see [74]. For graded preference voting see [75], [76], and [77]. ↩

References

1. Molteni, Megan. “Taiwan’s Digital Minister Knows How to Crush Covid-19: Trust.” *Wired*, September 30, 2020. https://www.wired.com/story/wired25-day3-audrey-tang-taiwan/. ↩

2. O'Brien, Thomas C., and Tom R. Tyler. "Authorities and Communities: Can Authorities Shape Cooperation with Communities on a Group Level?" *Psychology, Public Policy, and Law* 26, no. 1 (February 2020): 69–87. https://doi.org/10.1037/law0000202. ↩
3. Tsai, Lily L., Benjamin S. Morse, and Robert A. Blair. 2020. "Building Credibility and Cooperation in Low-Trust Settings: Persuasion and Source Accountability in Liberia During the 2014–2015 Ebola Crisis." *Comparative Political Studies* 53, no. 10-11 (September 2020): 1582–618. https://doi.org/10.1177/0010414019897698. ↩
4. Levi, Margaret, Audrey Sacks, and Tom Tyler. "Conceptualizing Legitimacy, Measuring Legitimating Beliefs." *American Behavioral Scientist* 53, no. 3 (November 2009): 354–75. ↩
5. Papada, Evie, David Altman, Fabio Angiolillo, Lisa Gastaldi, Tamara Köhler, Martin Lundstedt, Natalia Natsika, Marina Nord, Yuko Sato, Felix Wiebrecht, et al. "Defiance in the Face of Autocratization. Democracy Report 2023." V-Dem Working Paper, September 2023. https://doi.org/10.2139/ssrn.4560857. ↩
6. Rawls, John. "The Idea of Public Reason Revisited." *The University of Chicago Law Review* 64, no. 3 (Summer 1997): 765–807. https://doi.org/10.2307/1600311. ↩
7. Habermas, Jürgen. *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*, translated by Thomas McCarthy. Boston: Beacon Press, 1984. ↩
8. Habermas, Jürgen. *The Theory of Communicative Action, Volume 2: Reason and the Rationalization of Society*, translated by Thomas McCarthy. Boston: Beacon Press, 1987. ↩
9. Habermas, Jürgen. "Three Normative Models of Democracy." In *Constitutionalism and Democracy*, The International Library of Essays in Law and Legal Theory, edited by Richard Bellamy, 277–86. London: Routledge, 2017. ↩
10. Hill, T. *Autonomy and Self-Respect*. Cambridge: Cambridge University Press, 1991. https://doi.org/10.1017/CBO9780511609237. ↩
11. Hill, T. *Dignity and Practical Reason in in Kant's Moral Theory*. Ithaca: Cornell University Press, 1992. ↩
12. *In re Lawrence*, 190 P.3d 535, 560 (Cal. 2008). ↩
13. Argyle, Lisa P., Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. "Leveraging AI for Democratic Discourse: Chat Interventions Can Improve Online Political Conversations at Scale." *Proceedings of the National Academy of Sciences* 120, no. 41 (October 2023): e2311627120. https://doi.org/10.1073/pnas.2311627120. ↩

14. Mansbridge, Jane. “Rethinking Representation.” *American Political Science Review* 97, no. 4 (December 2003): 515–28. https://doi.org/10.1017/S0003055403000856. ↩
15. Lafont, Cristina. “Can Public Reason Be Inclusive?” In *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*, edited by Cristina Lafont, 191–218. Oxford: Oxford University Press, 2019. ↩
16. Gastil, John. *By Popular Demand: Revitalizing Representative Democracy through Deliberative Elections*. Oakland: University of California Press, 2000. ↩
17. Morrell, Michael E. “Deliberation, Democratic Decision-Making and Internal Political Efficacy.” *Political Behavior* 27, no. 1 (March 2005): 49–69. https://doi.org/10.1007/s11109-005-3076-7. ↩
18. Bächtiger, André, John S. Dryzek, Jane Mansbridge, and Mark E. Warren, eds. *The Oxford Handbook of Deliberative Democracy*. Oxford: Oxford University Press, 2018. ↩
19. Mansbridge, Jane. “A Minimalist Definition of Deliberation.” In *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies*, edited by Patrick Heller and Vijayendra Rao, 27–50. Washington, DC: World Bank Group, 2015. ↩
20. Dryzek, John S. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press, 2000. ↩
21. Karpowitz, Christopher F., and Tali Mendelberg. “An Experimental Approach to Citizen Deliberation.” In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Greene, James H. Kuklinski, and Arthur Lupia, 258–72. Cambridge: Cambridge University Press, 2011. ↩
22. Warren, Mark. 1992. “Democratic Theory and Self-Transformation.” *American Political Science Review* 86, no. 1 (March 1992): 8–23. https://doi.org/10.2307/1964012. ↩
23. Fishkin, James S. *The Voice of the People: Public Opinion and Democracy*. New Haven: Yale University Press, 1995. ↩
24. Chambers, Simone. *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse*. Ithaca: Cornell University Press, 1996. ↩
25. Mansbridge, Jane J. *Beyond Adversary Democracy*. Chicago: University of Chicago Press, 1983. ↩
26. Gutmann, Amy, and Dennis F. Thompson. “Reflections on Deliberative Democracy: When Theory Meets Practice.” In *The Oxford Handbook of Deliberative Democracy*, edited by André Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark E. Warren, 900–12. Oxford: Oxford University Press, 2018. ↩

27. Fishkin, James S., and Robert C. Luskin. “Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion.” *Acta Politica* 40, no. 3 (September 2005): 284–98. https://doi.org/10.1057/palgrave.ap.5500121. ↩
28. Gutmann, Amy, and Dennis F. Thompson. *Why Deliberative Democracy?* Princeton: Princeton University Press, 2004. ↩
29. Chambers, Simone. “The Philosophic Origins of Deliberative Ideals.” In *The Oxford Handbook of Deliberative Democracy*, edited by André Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark E. Warren, 55–69. Oxford: Oxford University Press, 2018. ↩
30. Polletta, Francesca, and Beth Gardner. “The Forms of Deliberative Communication.” In *The Oxford Handbook of Deliberative Democracy*, edited by André Bächtiger, John S. Dryzek, Jane Mansbridge, and Mark E. Warren, 70–85. Oxford: Oxford University Press, 2018. ↩
31. Lafont, Cristina. “Deliberation, Participation, and Democratic Legitimacy: Should Deliberative Mini-Publics Shape Public Policy?” *Journal of Political Philosophy* 23, no. 1 (March 2015): 40–63. https://doi.org/10.1111/jopp.12031. ↩
32. Thompson, Dennis F. “Deliberative Democratic Theory and Empirical Political Science.” *Annual Review of Political Science* 11, no. 1 (June 2008): 497–520. https://doi.org/10.1146/annurev.polisci.11.081306.070555. ↩
33. Cohen, Joshua. “Deliberation and Democratic Legitimacy.” In *The Good Polity: Normative Analysis of the State*, edited by Alan Hamlin and Philip Petit, 17–34. Oxford: Blackwell, 1989. ↩
34. Druckman, James N. “Political Preference Formation: Competition, Deliberation, and the (Ir)Relevance of Framing Effects.” *American Political Science Review* 98, no. 4 (November 2004): 671–86. https://doi.org/10.1017/S0003055404041413. ↩
35. Barabas, Jason. “How Deliberation Affects Policy Opinions.” *American Political Science Review* 98, no. 4 (November 2004): 687–701. https://doi.org/10.1017/S0003055404041425. ↩
36. Lafont, Cristina, ed. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford: Oxford University Press, 2019. ↩
37. Goldberg, Saskia. “Just Advisory and Maximally Representative: A Conjoint Experiment on Non-Participants’ Legitimacy Perceptions of Deliberative Forums.” *Journal of Deliberative Democracy* 17, no. 1 (April 2021), 56–75. https://doi.org/10.16997/jdd.973. ↩

38. Lafont, Cristina. “Can Democracy Be Deliberative & Participatory? The Democratic Case for Political Uses of Mini-Publics.” *Daedalus* 146, no. 3 (July 2017): 85–105. https://doi.org/10.1162/DAED_a_00449. ↩
39. Parkinson, John. “Legitimacy Problems in Deliberative Democracy.” *Political Studies* 51, no. 1 (March 2003): 180–96. https://doi.org/10.1111/1467-9248.00419. ↩
40. Lera, Sandro Claudio, Alex Pentland, and Didier Sornette. “Prediction and Prevention of Disproportionally Dominant Agents in Complex Networks.” *Proceedings of the National Academy of Sciences* 117, no. 44 (October 2020): 27090–95. https://doi.org/10.1073/pnas.2003632117. ↩
41. Braley, Alia, Gabriel S. Lenz, Dhaval Adjodah, Hossein Rahnama, and Alex Pentland. “Why Voters Who Value Democracy Participate in Democratic Backsliding.” *Nature Human Behaviour* 7, no. 8 (August 2023): 1282–93. https://doi.org/10.1038/s41562-023-01594-w. ↩
42. Horton, Chris. “The Simple but Ingenious System Taiwan Uses to Crowdsource Its Laws.” *MIT Technology Review*, August 21, 2018. https://www.technologyreview.com/2018/08/21/240284/the-simple-but-ingenious-system-taiwan-uses-to-crowdsource-its-laws/. ↩
43. The Computational Democracy Project. “Featured Case Studies.” Polis Knowledge Base, accessed: November 29, 2023. https://compdemocracy.org/Case-studies/. ↩
44. Anthropic. “Collective Constitutional AI: Aligning a Language Model with Public Input.” Published October 17, 2023. https://www.anthropic.com/index/collective-constitutional-ai-aligning-a-language-model-with-public-input. ↩
45. Miller, Carl. “Elon Musk Embraces Twitter’s Radical Crowdsourcing Experiment.” *Wired*, November 20, 2022. https://www.wired.com/story/elon-musk-embraces-twitters-radical-crowdsourcing-experiment/. ↩
46. Almaatouq, Abdullah, Alejandro Noriega-Campero, Abdulrahman Alotaibi, P. M. Krafft, Mehdi Moussaid, and Alex Pentland. “Adaptive Social Networks Promote the Wisdom of Crowds.” *Proceedings of the National Academy of Sciences* 117, no. 21 (May 2020): 11379–86. https://doi.org/10.1073/pnas.1917687117. ↩
47. Adjodah, Dhaval, Yan Leng, Shi Kai Chong, P. M. Krafft, Esteban Moro, and Alex Pentland. “Accuracy-Risk Trade-Off Due to Social Learning in Crowd-Sourced Financial Predictions.” *Entropy* 23, no. 7 (July 2021): 801. https://doi.org/10.3390/e23070801. ↩
48. Rajendra-Nicolucci, Chand, and Ethan Zuckerman. “An Illustrated Field Guide to Social Media.” Knight First Amendment Institute at Columbia University, May 14, 2021. http://knightcolumbia.org/blog/an-illustrated-field-guide-to-social-media. ↩

49. Krafft, P. M., Erez Shmueli, Thomas L. Griffiths, Joshua B. Tenenbaum, and Alex “Sandy” Pentland. “Bayesian Collective Learning Emerges from Heuristic Social Learning.” *Cognition* 212 (July 2021): 104469. https://doi.org/10.1016/j.cognition.2020.104469. ↩

50. Blair, Graeme, Jeremy M. Weinstein, Fotini Christia, Eric Arias, Emile Badran, Robert A. Blair, Ali Cheema, Ahsan Farooqui, Thiemo Fetzer, Guy Grossman, et al. “Community Policing Does Not Build Citizen Trust in Police or Reduce Crime in the Global South.” *Science* 374, no. 6571 (November 2021): 1046–7. https://doi.org/10.1126/science.abd3446. ↩

51. Bakker, Michiel, Martin Chadwick, Hannah Sheahan, Michael Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, Amelia Glaese, John Aslanides, Matt Botvinick, et al. “Fine-Tuning Language Models to Find Agreement among Humans with Diverse Preferences.” *Advances in Neural Information Processing Systems 35*, edited by S. Koyejo, S. Mohammed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, 38176–89. Red Hook: Curran Associates, 2023. https://proceedings.neurips.cc/paper_files/paper/2022/hash/f978c8f3b5f399cae464e85f72e28503-Abstract-Conference.html. ↩

52. Office of Intelligence and Analysis. “Homeland Threat Assessment 2024.” US Department of Homeland Security, 2024. https://www.dhs.gov/sites/default/files/2023-09/23_0913_ia_23-333-ia_u_homeland-threat-assessment-2024_508C_V6_13Sep23.pdf?utm_campaign=wp_the_cybersecurity_202&utm_medium=email&utm_source=newsletter&wpisrc=nl_cybersecurity202. ↩

53. Funk, Allie, Adrian Shahbaz, and Kian Vesteinsson. “The Repressive Power of Artificial Intelligence.” Freedom House, 2023. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence. ↩

54. Bond, Shannon. “Fake Viral Images of an Explosion at the Pentagon Were Probably Created by AI.” National Public Radio, Untangling Disinformation, May 22, 2023. https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai. ↩

55. Twomey, John Joseph, Conor Linehan, and Gillian Murphy. 2023. “Deepfakes in Warfare: New Concerns Emerge from Their Use around the Russian Invasion of Ukraine.” *The Conversation*, October 26, 2023. http://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393. ↩

56. Klepper, David. 2023. “Fake Babies, Real Horror: Deepfakes from the Gaza War Increase Fears about AI’s Power to Mislead.” Associated Press, November 28, 2023. https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47. ↩

57. Banias, M. J. “Inside CounterCloud: A Fully Autonomous AI Disinformation System.” *The Debrief*, August 16, 2023. https://thedebrief.org/countercloud-ai-disinformation/. ↩

58. Tsai, Lily L. *When People Want Punishment: Retributive Justice and the Puzzle of Authoritarian Popularity*. Cambridge: Cambridge University Press, 2021. ↩

59. Blair, Robert A., Benjamin S. Morse, and Lily L. Tsai. “Public Health and Public Trust: Survey Evidence from the Ebola Virus Disease Epidemic in Liberia.” *Social Science & Medicine* 172, no. 1 (January 2017): 89–97. https://doi.org/10.1016/j.socscimed.2016.11.016. ↩

60. “MIT Scholars Awarded Seed Grants to Probe the Social Implications of Generative AI,” MIT News | Massachusetts Institute of Technology, September 18, 2023, https://news.mit.edu/2023/mit-scholars-awarded-seed-grants-generative-ai-0918. ↩

61. Singh, Spandana. “Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content.” Open Technology Institute, July 22, 2019. http://newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/. ↩

62. Kojaku, S., R. Mahari, S. Lera, E. Moro, A. Pentland, and Y. Y. Ahn. “Uncovering the Universal Dynamics of Citation Systems: From Science of Science to Law of Law and Patterns of Patents.” Presented at the International School and Conference on Network Science, Vienna, Austria, July 2023. ↩

63. Motoki, Fabio, Valdemar Pinho Neto, and Victor Rodrigues. “More Human Than Human: Measuring ChatGPT Political Bias.” *Public Choice* (August 2023): 1–21. ↩

64. Habermas, Jürgen. *Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des Demokratischen Rechtsstaats*. Frankfurt: Suhrkamp, 1992. ↩

65. Dickson, Eric S., Catherine Hafer, and Dimitri Landa. “Cognition and Strategy: A Deliberation Experiment.” *The Journal of Politics* 70, no. 4 (October 2008): 974–89. https://doi.org/10.1017/S0022381608081000. ↩

66. Steenbergen, Marco R., André Bächtiger, Markus Spörndli, and Jürg Steiner. “Measuring Political Deliberation: A Discourse Quality Index.” *Comparative European Politics* 1, no. 1 (March 2003): 21–48. https://doi.org/10.1057/palgrave.cep.6110002. ↩

67. De Vries, Raymond, Aimee Stanczyk, Ian F. Wall, Rebecca Uhlmann, Laura J. Damschroder, and Scott Y. Kim. “Assessing the Quality of Democratic Deliberation: A Case Study of Public Deliberation on the Ethics of Surrogate Consent for Research.” *Social Science & Medicine* 70, no. 12 (June 2010): 1896–903. https://doi.org/10.1016/j.socscimed.2010.02.031. ↩

68. Stromer-Galley, Jennifer. “Measuring Deliberation’s Content: A Coding Scheme.” *Journal of Deliberative Democracy* 3, no. 1 (July 2007): 12. https://doi.org/10.16997/jdd.50. ↩

69. Edwards, Peter B., Richard Hindmarsh, Holly Mercer, Meghan Bond, and Angela Rowland. “A Three-Stage Evaluation of a Deliberative Event on Climate Change and Transforming Energy.” *Journal of Public Deliberation* 4, no. 1 (January 2008): 6. https://doi.org/10.16997/jdd.65. ↩

70. Parkinson, John, and Jane Mansbridge, eds. *Deliberative Systems: Deliberative Democracy at the Large Scale*, Theories of Institutional Design. Cambridge: Cambridge University Press, 2012. ↩

71. Mockler, Patricia. “Measuring the Inclusiveness of Deliberation: Structural Inequality and the Discourse Quality Index.” *Comparative European Politics* 20, no. 1 (February 2022): 53–72. https://doi.org/10.1057/s41295-021-00262-5. ↩

72. Rowe, Gene, and Lynn J. Frewer. “Public Participation Methods: A Framework for Evaluation.” *Science, Technology, & Human Values* 25, no. 1 (Winter 2000): 3–29. https://doi.org/10.1177/016224390002500101. ↩

73. Timotijevic, Lada, and Monique Maria Raats. 2007. “Evaluation of Two Methods of Deliberative Participation of Older People in Food-Policy Development.” *Health Policy* 82, no. 3 (August 2007): 302–19. https://doi.org/10.1016/j.healthpol.2006.09.010. ↩

74. Posner, Eric A., and E. Glen Weyl. “Voting Squared: Quadratic Voting in Democratic Politics.” *Vanderbilt Law Review* 68, no. 2 (March 2015): 441–500. ↩

75. Camps, Rosa, Xavier Mora, and Laia Saumell. “A Continuous Rating Method for Preferential Voting. The Incomplete Case.” *Social Choice and Welfare* 40, no. 4 (April 2013): 1111–42. https://doi.org/10.1007/s00355-012-0663-5. ↩

76. Kacprzyk, Janusz, and Mario Fedrizzi. “Consensus Degrees Under Fuzziness via Ordered Weighted Average (OWA) Operators.” In *Fuzzy Logic and Its Applications to Engineering, Information Sciences, and Intelligent Systems*. Theory and Decision Library Series D: System Theory, Knowledge Engineering and Problem Solving 16, edited by Z. Bien, and K. C. Min, 447-53. Dordrecht: Kluwer Academic Publishers, 1995. https://doi.org/10.1007/978-94-009-0125-4_44. ↩

77. Yager, Ronald R., and Janusz Kacprzyk, eds. *The Ordered Weighted Averaging Operators: Theory and Applications*. New York: Springer, 2012. ↩