AI in Deliberative Democracy: Empirical Insights on the legitimacy of AI-supported deliberation platforms

Alina Östling, ORCID 0000-0002-7702-5957
Visvaldis Valtenbergs, ORCID 0000-0001-6276-7475
Velta Skolmeistere, ORCID 0000-0002-3704-4547

"Trying to have a political conversation on Facebook is like trying to have a political conversation in a nightclub."[1] Taiwan's Minister of Digital Affairs, Audrey Tang

Abstract

This chapter presents an empirical examination of deliberative platforms to understand how Artificial Intelligence (AI) facilitates or challenges democratic legitimacy. This review contributes with an analysis of the existing deliberative platforms powered by AI, offering a nuanced understanding of the potential and risks associated with employing AI to augment public discourse and decision quality in the digital age.

Introduction: The novelty of the AI-supported platforms

According to an assessment by the Varieties of Democracy (V-Dem), deliberative democracy has been steadily deteriorating across the globe, with a marked decline observed over the past decade (Papada et al. 2023). Concurrently, online platforms designed to facilitate deliberative engagement have proliferated in both Europe and the United States. However, theoretical and empirical studies have voiced significant critiques regarding the capacity of these online platforms, particularly social media platforms to foster deliberative norms and enhance deliberative processes. Critics have identified that online environments often encourage uncivil discourse, manifesting in forms such as flaming, sexism, racism, and hate speech (Buchstein 1996; Suler 2004). Additionally, social network discussions have been criticised for engendering irrational discourses—including climate change denial and conspiracy theories—that undermine informed public debate (Davis 2021). These discussions may even contribute to the propagation of low-quality discourse, which skews public perceptions on critical issues (Anderson et al. 2014, 2018). Many of these deficiencies can be attributed to the inherent limitations in human moderation and the management of online interactions, underscoring the pressing need for alternatives. By leveraging advanced algorithms and machine learning techniques, AI and machine learning in theory can facilitate a more constructive and rational dialogue, effectively addressing the challenges posed by the current landscape of online deliberation. Some researchers and practitioners have actually started to examine the potential of the AI to solve some of the problems associated with online deliberation (Rodríguez-Ruiz et al. 2020; Stoll et al. 2020; Wojcieszak et al. 2021; Friess et al. 2022; Carstens, Friess, 2023). Several online participation platforms have also started using or are planning to use AI to facilitate the development of mutual respect

[1] Molteni, M. (30.09.2020). "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/

and understanding during deliberations, to summarise citizen opinions, or to enable participants to converge in their views (Tsai et al. 2024, Deen et al. 2012). One of the most publicised use-cases is *vTaiwan* - an online platform in Taiwan launched in 2015. By leveraging *pol.is*, an AI tool that aggregates and organises public opinions in real time, *vTaiwan* allows users to see how their views align or diverge from those of others. This interactive feature enables participants to navigate complex issues collaboratively, fostering a sense of community and encouraging deeper engagement in the policymaking process.

vTaiwan and other platforms seem to assume that technical design features can support deliberative discussions among participants (Deen et al. 2012). Certain types of architecture are believed to nudge people toward better behaviours and result in improved quality of deliberation. Some of these features are not completely new but, without the support of advanced technology tools (including AI) they have not been easy to implement, e.g. requiring intensive and costly use of human resources to manage a large number of participants (Deen et al. 2012). AI-assisted online deliberative platforms are addressing some of these limitations and claim to achieve better deliberation faster and at a larger scale, with less human bias and lower costs (Tsai et al. 2024). A recent review of these platforms carried out by the MIT also argues that these technologies could help address democratic challenges such as polarisation, mistrust, and democratic backsliding (Tsai et al. 2024). Given the pace of the development and the potential transformation that these platforms bring about, it's important to explore what implications they have for deliberative democracy; and this is what we aim for in this chapter.

Methodology

This chapter aims to make a link between normative democratic theory and the empirical study of AI-supported deliberation platforms. Our research design is based on the "inductive political theory", coined by Hélène Landemore (2020).[2] This method builds on the generalisation and deeper examination of initiatives tested on the ground by activists, politicians or other actors (Landemore 2020 in Fischli and Muldoon 2024). Further interpreted by Fischli and Muldoon (2024), inductive political theory is a method to learn from an exploration of democratic experiments aiming at developing normative concepts and principles. The method stands in contrast to the deductive forms of political theory, which tend to take theoretical principles as given and apply them to real-world contexts (Fischli and Muldoon 2024).

Our initial sample was a list of deliberative platforms compiled by Tsai et al. (2024) (see Annex 2 in the report "Generative AI for Pro-Democracy Platforms."), enumerated in Annex 1 of this chapter. The 27 platforms on the list facilitate deliberations that involve lay citizens online, have public institutions among their "clients", and include discursive elements (Tsai et al 2024).

The chapter draws on existing knowledge and research related to the AI-supported deliberation platforms, using the information about the platforms publicly available on the Internet, academic articles, and a survey of the organisations running the platforms. Although several of these platforms have been discussed in

[2] The term was coined by Landemore according to this article. (...) in her 2020 book and in a 2021 article in Raisons Politiques, (Landemore) enunciated the idea of "inductive political theory."
https://link.springer.com/article/10.1057/s41296-023-00615-7

previous literature, (to our knowledge) they have not yet been systematically examined in relation to the democratic legitimacy dimensions.

Our research commenced with an analysis of the information related to our research questions (see below) available on the landing pages of the respective platforms. To further verify and enrich our data, we conducted an additional validation by distributing a concise email survey to representatives of these platforms. The primary objective of the survey was to authenticate the insights obtained from our desk research, while gaining a deeper comprehension of AI's role in enhancing the legitimacy of deliberative processes. The survey queried whether the platforms employed AI or machine learning methodologies. For platforms affirming the use of AI, we requested detailed information about specific AI functionalities utilised to facilitate deliberative processes. Respondents were provided with a list of potential AI functionalities to select from, along with the option to elaborate on additional functionalities. In addition our survey sought to determine whether and how these platforms ensure their AI tools are auditable and transparent. Finally, we invited the platforms to evaluate the impact of AI on making deliberations more inclusive, transparent, accountable, and balanced.[3]

We initially assumed that all platforms in our survey sample employed some form of AI or machine learning. However, upon administering a control question—"Do you use artificial intelligence (AI) or machine learning in any capacity on your platform?"—we discovered that only fifty percent of the responding platforms (6 out of 12) confirmed their use of AI technologies. As a result, the discussion of our survey findings mainly focuses on these six platforms. From our survey results, it is noteworthy that several platforms, which currently do not integrate AI, are actively exploring machine learning and generative AI in various experimental and precautionary ways. These exploratory efforts are also discussed in this chapter.

Given that we do not examine deliberative platforms from a technical point of view, we used the term AI interchangeably with machine learning. In terms of the type of AI used, recent findings suggest that many deliberative platforms have incorporated natural language processing (NLP)[4] and sentiment analysis, and only very few have begun integrating generative AI (Tsai et al. 2024). However, in our study, we did not delve deeper into the technicalities and implications of the kind of AI used by the platforms.

The definition and operationalisation of key concepts

The definition of deliberation and legitimacy

In the context of our research we adopt a general definition of deliberation as an exchange of reasons, arguments, and justifications (Landemore 2023). From a normative perspective, we assume that deliberation has to have a certain level of democratic legitimacy.

[3] The E-mail survey template is available at this link: https://drive.google.com/file/d/1iWPdY4HS9C9g7CbBv9Rdf7fOO8dhgXPO/view?usp=sharing

[4] Natural language processing (NLP) is an interdisciplinary subfield of computer science and artificial intelligence. https://en.wikipedia.org/wiki/Natural_language_processing

According to Scharpf (1997), democratic legitimacy refers to the inputs and outputs of a political system. On the *input* side, legitimacy requires mechanisms to connect political decisions with the preferences of citizens (Scharpf 1997). In order to achieve a higher quality of input legitimacy, deliberation should be inclusive and representative of the population affected by a (political) output (Barber, 1984; Lieberherr et al., 2012). Inclusiveness means that the platforms are open to all individuals and to the organised civil society, and involve them in such a way as to ensure balance in their representation (Schmidt and Wood 2019). This can presume either a representative sample of the population, or representation of groups of the population concerned by the issues discussed. Legitimate deliberation also means that all members of the demos participate on equal grounds and that all are considered as equals (Landemore 2023).

Output legitimacy is mainly about performance in terms of policy effectiveness or other results (Scharpf 1999; Majone 1998; Bellamy 2010; in Schmidt and Wood 2019), assuming that institutions are legitimised based on their effectiveness (Lieberherr and Thomann, 2018). Output legitimacy also envisions a certain level of transparency, such as access to information about the resulting outputs, e.g. decisions or findings (Stephenson 2023).

Throughput legitimacy, coined by Vivien Schmidt, complements Scharpf's dimensions by highlighting the process between input and output (Schmidt 2013). The openness of the governance processes (Schmidt, 2013; Schmidt and Wood, 2019) at the core of throughput legitimacy. When opinions and options emerging through deliberation are closed to scrutiny, both input and outputs of the deliberative exercise can be questioned (Schmidt 2013). To this, we would also add the importance of the openness of the deliberation platforms themselves, as well the transparency of the technologies operating the platforms. In the context of digital platforms, transparency could be (partly) fulfilled by e.g. the use of open source to allow platform design and technologies to be scrutinised by third parties. In addition to openness, accountability of the processes, requiring an independent oversight mechanism, makes the deliberation legitimate (Schmidt, 2013; Schmidt and Wood, 2019).

Moreover, to achieve a high quality of deliberation process, participants should listen to each other, justify their positions, show respect for counterarguments, and be willing to re-assess their initial preferences through discussion (Coppedge et al. 2024, Steenbergen et al. 2003).[5] According to deliberative theorists this will lead to well-informed preferences and *more legitimate decisions* (Fishkin 1995, Cohen 1989 in Steenbergen et al. 2003), in contrast to a mere aggregation of existing preferences. An ideal deliberative process is one in which public reasoning, focused on the common good, motivates political decisions. The dialogue among informed and competent participants who are open to persuasion should be respectful during the whole process, from preference formation to final decision (Papada et al. 2023, p. 55). Balanced deliberation stands in contrast to domination of the debate and group polarisation, and can be supported with (human or AI-supported) moderation, and an adequate platform design (Coppedge et al. 2024, Steenbergen et al. 2003).

In our chapter, we map and discuss the deliberative platforms identified, focusing on the role of AI as a potential enabling force of democratic legitimacy. However, we mainly focus on the dimensions of input and throughput but do not examine the role of AI when it comes to output legitimacy. This is due to several

[5] These criteria follow the principles of Habermas' discourse ethics, as outlined in Steenbergen et al. (2003).

reasons. For one it is quite difficult to isolate the impact of the AI on the deliberative outputs from other contextual factors (such as societal, political and organisational factors). It has been argued in e-participation research that the outcomes of e-participation relate more to these factors rather than technical aspects (e.g, Chadwick 2011; Zheng et al. 2014; Jho & Song 2015). In addition, while researching the different participation and deliberation platforms, we did not find convincing evidence for how AI could have affected the outputs of the deliberative process. This could, however, be an interesting avenue of future research, especially if the definition of output legitimacy is broadened beyond efficiency e.g. to cover improved cognition (see e.g. Dennis Friess and Carina Weinmann's chapter in this book). Hence, our research scope focuses more on the role of AI in affecting inclusiveness and processes of deliberation:

- *RQ1. Does AI affect the inclusiveness of deliberation[6] (input legitimacy)?*
- *RQ2: Does AI affect the process of deliberation (throughput legitimacy)?*

Overview of AI functionalities used by deliberation platforms

The surveyed platforms reported the integration of various AI functionalities to support their operations. A common thread is the emphasis on automated reporting, which emerged as the most frequently utilised feature among four platforms. This reflects a growing recognition of the need for efficiency in information dissemination and decision-making processes. Text analysis using Natural Language Processing (NLP) and translation and linguistic support were each integrated by three platforms. In addition, *Stanford Deliberate* detailed the implementation of diverse AI features, such as text analysis, translation, and summarization, along with innovative applications of NLP for text quality estimation and transcription, differing slightly from the questionnaire's scope. This indicates a desire to facilitate clearer communication among diverse participants, reducing language barriers and enhancing the accessibility of discussions. By employing these AI tools, platforms aim to create more inclusive environments where varying perspectives can be shared and understood, ultimately enriching the deliberative process.

The platforms also demonstrate a commitment to deploying advanced AI functionalities, including recommendation systems, automated moderation, and real-time deliberation quality measurements. These advanced tools highlight a trend towards creating smarter, adaptive environments where discussions can be guided more effectively, enabling participants to focus on substantive issues.

Notably, no platform reported employing AI for fact-checking purposes. This raises critical questions about the integrity and credibility of information presented during deliberations. The lack of focus on verification methods suggests a potential vulnerability in these digital participation spaces, indicating that while AI enhances engagement, further efforts may be necessary to ensure the reliability of discourse content.

Certain platforms, including *Stanford Deliberate* and *Democracy developers,* utilised external AI models. *Democracy Developers* specifically mentioned their use of *GloVe* for information retrieval, stating that user

[6] Inclusive deliberation: democratic deliberation that is inclusive of all members of the demos on equal grounds (Landemore 2023).

queries are matched to "similar" questions based on *GloVe'*s pre-trained word vectors. Additionally, *Your Priorities* incorporates an AI agent workflow through its *Policy Synth* toolkit, facilitating an agentic AI process synergistically blended with deliberation. *Policy Synth* constitutes an innovative toolkit that enhances problem-solving by automating the identification and prioritisation of constituent problems (Bjarnason, Gambrell, Lanthier-Welch 2024).

Specialised functionalities, such as *Discuto's* dynamic discussion capabilities and the "consensus meter," illustrate a growing recognition of the need for interactive tools that foster real-time adaptability and consensus-building within groups. By allowing documents to evolve concurrently with discussions, these platforms empower participants to engage more deeply with content, actively seeking areas of agreement and understanding throughout the deliberative process.

In summary, while the surveyed platforms reveal a range of AI functionalities that facilitate participation and enhance deliberation, the overarching themes highlight both promising advancements and areas for improvement. The integration of AI tools reflects an adaptive response to the challenges of diverse participation and complex deliberative spaces. However, the lack of prioritisation of fact-checking mechanisms and the need for human oversight remain critical considerations. For a detailed overview of AI functionalities used by surveyed deliberation platforms, consult Annex 2.

The opportunities of AI in deliberation

Input legitimacy

When examining input legitimacy, we mainly focused on how AI can support the inclusiveness of the deliberation process. Both in the literature and in our survey of the platforms, we found evidence of related AI-features such as multilingual support and delegation of votes to someone more knowledgeable on the topic. In our survey, we specifically asked platforms to agree or disagree with the statement "Using AI in our platform makes the deliberation more inclusive". Out of the four responses received to that question, one respondent tended to disagree, while three strongly agreed with this statement. The other respondent who strongly agreed, added a disclaimer that their platform is not implementing inclusiveness measures yet, but only planning to. None of the respondents provided any sources of evidence to underpin their answers.[7]

Multilingual support

Several platforms use AI to provide translation in order to facilitate communication between users - *Assembl, Consider.it, CoUrbanize*, and *Your Priorities* (Tsai et al. 2024). For example, *Your Priorities* uses open APIs to provide real-time translation and transcription of videos, by leveraging integrated *Google* machine translation (Solonian Democracy Institute 2024).[8] *Assembl* has a chatbot feature, which uses large

[7] In the survey, we asked for evidence such as external evaluations, academic articles etc.

[8] Fuller, R. H., Jakovljev, M. (2024). *SDI Digital Democracy Report*. Solonian Democracy Institute, 25. and *YourPriorities*, https://citizens.is/your-priorities-features-overview/

language models (*LLMs*)[9] to generate text-based output, in order to help citizens write comments (Tsai et al. 2024). These types of multilingual support features could be useful in increasing the involvement of users with different levels of linguistic skills, e.g. people that are not used to communicating in writing or who are not native speakers. One of the survey respondents stressed that "multilingual deliberations increases inclusivity, e.g. in border regions and where, for example due to migration, there is no lingua franca equally adopted by everyone in the area."

Trust networks: voters as vectors

A distinctive feature that might lead to increased input legitimacy is being developed by the platform *Ethelo*.[10] Its decision algorithm incorporates the ability to support "Liquid Trust", which is a feature building on "liquid democracy", a voting system that enables participants to either vote on issues directly or to delegate their voting power to someone with more knowledge in a particular area. This delegation of votes, one might argue, could enable people to participate more fully in political decisions, given that they can leverage on more experience and skilled proxies (Green-Armytage 2014). The voter can assign "trust weights" to several delegates that can vote instead of her/him depending on the issue or criteria (Ethelo 2021). Issue-competent delegates can presumably make better decisions compared to representatives who decide on all issues regardless of expertise (Alger 2006). In theory, this means that trusted delegates - who are presumably more knowledgeable and experienced on specific issues compared to the person that delegates the vote - can better contribute to debate and decisions (Ethelo 2018).

The integration of liquid democracy and trust vectors within AI-enabled systems may provide an impulse for more inclusive participation by accommodating varying levels of interest, expertise, and time commitment, thus enabling a broader spectrum of society to engage meaningfully with policy issues. Trust vectors may also help to identify credible sources and trustworthy peers, thereby mitigating the risk of misinformation and building confidence in the deliberative process. However, in order for this mechanism to work properly and contribute to the legitimacy of the input, the person selecting a trusted proxy should be applying reasoned thinking and relevant criteria when delegating their power.

Throughput legitimacy

Balanced deliberation

When we asked "To what extent do you agree with this statement: "Using AI in our platform makes the deliberation more balanced?" In our survey, only one respondent strongly agreed (score 5), while three stayed neutral (score 3). The respondent who strongly agreed with the statement, mentioned that AI helps the platform identify the "topics that need most care", while another respondent said that they use deterministic approaches (process design and preference aggregation algorithms) to encourage constructive behaviour among users but without using AI. One of the respondents simply commented that balanced deliberation is not the reason for their use of AI.

[9] LLM is a subtype of generative AI (Tsai et al. 2024).
[10] The *Ethelo* platform does not yet include Liquid Trust-interfaces, but the data architecture is already in place. Source: https://blog.ethelo.org/use-trust-networks-next-level-liquid-democracy

Mutual respect

According to a similar empirical study by Tsai et al. (2024), some platforms have already begun to integrate NLP and sentiment analysis to flag toxic speech (*Assembl, Stanford Online Deliberation Platform (SODP),* and *Your Priorities*). For example, the AI of *Your Priorities* scans content to detect toxicity; if something toxic makes it onto the platform, community administrators get a notification right away (Citizens, n.d.). The *SODP* uses several approaches to support balanced deliberation, including an automated moderator (based on an algorithm called "Alice") to report instances of toxic speech during small-group discussions. When toxic behaviour is detected, or when a participant is reported, the automated moderator asks the group to validate this and takes action accordingly. *SODP* is also facilitating small group discussions by using a speaker queue and nudging participants that have not spoken much (Stanford Online Deliberation, n.d.).

Consensus-building

One of the prominent platforms when it comes to consensus building is *Pol.is*. According to *Pol.is'* cofounder Colin Megill, with new computational methods it is possible to find areas of consensus, to contrast division (Leonard, 2020). Pol.is is designed in a way that assures that "people can only add to, but not subtract or detract from the conversation" (Leonard, 2020). These claims are supported by research carried out by Tsai et. al (2024) at MIT, which considers that *Pol.is'* design features could have a significant impact on decreasing polarisation.

Basically, *Pol.is* is using complex algorithms to identify patterns of agreement or disagreement among users, and generate consensus-based recommendations on complex and disputed issues. It has also developed an *upvote/downvote mechanism,* which creates a citation network (similar to citation networks used in scientific papers, patent applications, and legal decisions). The upvotes and downvotes are analogous to citations. The network of citations drives citizen interaction on the platform. "The process of surveying comments made by others for upvoting and downvoting forces people to learn about others' opinions" (Tsai et al., 2024).

In terms of the technical elements, *Pol.is* is using a self-learning algorithm, a "neural network", similar to *Netflix* or *Spotify* recommendation engines. The participant can see an opinion on a consultation topic and can react to it (agree, disagree, neutral), or propose an alternative opinion. He/she is restricted by the limited number of characters and the recommendation to produce only one idea per contribution. Eventually, the algorithm associates the participant with an opinion group adopting similar positions (Duval, 2018).

Several other platforms use algorithms in order to facilitate civility and consensus. *Ask Parliament, LiquidFeedback* and *Your Priorities* uprank comments with the most support, while *Make.org* upranks comments with the most positive sentiment (Tsai et al. 2024). *Assembl* uses AI to suggest actions to users, e.g. to provide a reason for their comment, give an example, or state their comment in more general terms (Tsai et al. 2024).

Ethelo is another platform with a complex algorithm targeting consensus. It is designed to find outcomes with high average support and low variance in satisfaction among participants - "a form of "optimized consensus" for large groups" (Ethelo, n.d.). This can complement processes such as quadratic voting, conjoint analysis, the Monte Carlo Method and decision trees. The algorithm combines multiple decision

parameters such as options, issues, criteria, constraints, influence, and fairness to develop a space of potential scenarios. Those scenarios are then evaluated and ranked according to the aggregated analysis of participants, prioritising fairness in the distribution of satisfaction (Ethelo, n.d.). Aiming at "optimized consensus", the *Ethelo* algorithm searches for the outcome that maximises average support, while minimising polarisation. This outcome will often have approval levels above 90%, and is allegedly able to capture compromises that are generally not surfaced through traditional democratic decision-making processes (Ethelo, 2021).

Categorisation and visualisation

Several platforms - including *Assembl*, *Go Vocal* and *Pol.is* - are using machine learning to produce graphical representations of topics and sentiments during deliberations (Tsai et al. 2024). The platform *Go Vocal/CitzenLab* also to provide insights into community feedback and engagement patterns (GoVocal, n.d. a) and uses natural language processing (NLP) to automatically group ideas around keywords and physical locations (Fuller & Jakovljev, 2024). Similarly, *CitizenLab* uses AI to consolidate large numbers of proposals into manageable documents. To ensure transparency, the summarised content contains links to the original proposals. This allows public officials to understand geographic differences in voter preferences (Fuller & Jakovljev, 2024). On the same line, Pol.is uses machine learning algorithms to analyse and visualise user input. *Pol.is* also allows participants to mouse over the visualisation to view the most popular reasons given in an opinion area (Tsai et al. 2024).

Your Priorities, in addition to ranking and organising submissions based on user feedback, also use *GPT-4* AI Assistant, to help citizens in navigating and understanding extensive projects, and assists governments in searching and summarising data (Citizens, n.d.). Other platforms, including A*ssembl, Decidim and LiquidFeedback*, also suggest clustering of proposals by common topics and propose keywords in comments. *LiquidFeedback* also uses AI to help users to avoid duplicates.[11] However, all these platforms preserve human agency by allowing human confirmation of the AI-based categorisation (Tsai et al. 2024).

One of the platforms, *Ethelo*, underlines that they do not see AI-tools approaching human capabilities in this area anytime soon. However, AI facilitates categorisation, especially once a sample set of comments have already been themed and coded. They believe that "Applied within a proper framework alongside human analysis there is an enormous opportunity for AI to speed the analysis and identify larger patterns across very large datasets." (Ethelo, 2021)

Transparency and explainable AI

According to our desk research about half (13 of 27) of the studied platforms use or plan to use open source technology.[12] *Ask Parliament/Democracy Developers* claims to have the first open source online debate software,[13] while Democracia en Red running Crowd Law Making developed their open source platform

[11] E-mail survey.

[12] Please note that there might be more platforms that use open source. We did not specifically focus on this dimension but took it into account when looking for information on the platforms' websites.

[13] They point to this reference for details: https://www.bluenove.com/en/blog/leader-collective-intelligence-acquisition-succeed-together/

in-house and actively promote the use of En Código para el Desarrollo, an initiative of the Inter-American Development Bank (Banco Interamericano de Desarrollo) that provides a repository of open source applications, algorithms, spreadsheets, and other useful digital tools to address specific challenges (Devenyi, Di Giacomo, O'Donohoe, 2021).

In our survey, four platforms (*LiquidFeedback, Your Priorities/Citizens Foundations, Ask Parliament/Democracy Developers, Go Vocal/CitzenLab*) stated that they have mechanisms in place to make the AI used on their platform auditable and explainable. *LiquidFeedback* reported that their "AI policy is to avoid any black box-like behaviour that could undermine verifiability or create actual or perceived bias." However, it's unclear what their policy embodies given that they did not share any policy document. *LiquidFeedback* also provided a few use cases that allegedly describe that the AI used is "inherently audible and explainable". For instance, they argue that their translation service of user generated content (aimed to facilitate multilingual deliberation) has less bias than a human translation, and can be reviewed by any user. However, they do not point to any information that is openly shared by the AI system, or the data used to train the system.

Ask Parliament/Democracy Developers use a public model (from *GloVe*) and their own open-source software (GitHub, n.d.). Although they have not specifically worked on explainability of their AI, they state that it has no hidden code or model; and the model is not currently (re)trained on their own questions. *Go Vocal/CitzenLab*, which also stated that they have mechanisms in place to make the AI used on their platform auditable and explainable, does not provide any information about the source code used to train and run the system. However, they mention that their AI input analysis assistant, which summarises the deliberation inputs, always provides references to the original statements from users to ensure that the quality can be fact-checked; as well as accuracy and confidence scores of the analysis (GoVocal, n.d. b). *Citizens Foundations* - which runs the open source *YourPriorities* platform - explains how some of the key technical components, including AI agents, *Elo* scoring, pairwise voting, and genetic algorithms, function within the platform in a scientific article, which is freely available on the web (Bjarnason, Gambrell, & Lanthier-Welch, 2024).

In summary, while all four surveyed platforms underscore their commitment to auditability and explainability of AI, the extent and nature of their efforts vary. Some platforms highlight user-facing transparency features, whereas others focus on the use of open-source models. *YourPriorities* distinguishes itself by providing detailed documentation on their AI systems' functionality, setting a benchmark for openness in AI operations. Collectively, these findings illustrate a diverse landscape of AI transparency initiatives across deliberative platforms, with varying levels of disclosure and emphasis on explainability.

The risks of AI-powered deliberation platforms

AI does not only bring new opportunities but also new risks for deliberation, although the latter are not discussed as much in the literature or by the platforms themselves. For instance, *Dialogue/Delib* adopts a precautionary approach, refraining from using AI or machine learning due to concerns about their potential impact on political decisions. They emphasise the need to ensure that any technology implemented aids public sector decision-making and enhances data processing without causing distortions: "we wish to be absolutely certain that whatever is brought in would aid public sector decision-making and/or help to speed

up the processing of data, rather than potentially distorting it."[14] Beyond these potential shortcomings, the risks associated with AI include limitations on agency, openness, transparency, and auditability, as well as challenges related to privacy and anonymity.

Tsai et al. underline that a democratic challenge to deliberative platforms is that people might be inclined to agree with AI-generated consensus statements not because they genuinely support these but because they are written in a more literate or agreeable way. The efficiency of AI might take away the creativity and openness to innovative ideas, as well as cast doubts about both the platform and process if participants feel that AI has undue influence (Tsai et al. 2024). Excessive agreement and consensus, especially if orchestrated by opaque AI mechanisms, can challenge the legitimacy of the deliberation. Moats and Tseng (2023), who have studied how *vTaiwan* (implemented by *Pol.is)*, argue that the platform's focus on votes at the expense of texts, and the display of distinct (and perhaps fewer) groups may lead to a less lively, pluralistic and nuanced conversation (Moats and Tseng 2023, p. 994). This also poses a risk to minority interests and views (Tsai et al. 2024).

Tsai et al. (2024) also assert that the principle of agency could be threatened when (untransparent) AI are making certain decisions on behalf of users, and downgrade certain opinions, e.g. due to inherent model biases (depending on training data) or manipulation by external entities, especially given the present risk of exploitation of AI systems by bad actors who can skew discussions or spread misinformation on deliberation platforms. Some platforms are of a similar opinion, e.g. *LiquidFeedback* abstain from using AI for moderation, as they feel this would undermine verifiability and trust of their users[15], and Wietse Van Ransbeeck, the CEO and Co-Founder of *Go Vocal*, states in a recent Forbes article that it's a challenge to maintain transparency and trust when machines are involved in the conversation (Van Ransbeeck, 2024). *Go Vocal/CitzenLab,* when answering our survey, also underlined that "(p)articipant agency is a rightful concern" on the deliberation platforms using AI.[16]

Moreover, AI-supported transparency is not a question at the core of the platforms that responded to our survey. Three respondents were neutral (neither agreed or disagreed), and one tended to disagree with the statement "Using AI in our platform makes the deliberation more transparent and accountable to the general public?". Moreover, the respondents did not give any relevant examples of how AI can support transparency on their platform in the comment section. In contrast, *Go Vocal/CitzenLab* underlined that their current approaches will probably not increase accountability: "rather the opposite when AI gets to summarise multiple perspectives into new statements".[17]

The lack of algorithm transparency is a significant challenge to deliberative platforms, e.g. in the ranking and presentation of comments, as well of proposed consensus areas, where decisions made by platform developers are not always clear or explained to the end-users (Tsai et al. 2024). For instance, generative AI can today identify likely areas of consensus even before deliberation starts, which could help to minimise disagreement and speed up the whole process. However, if the participants do not have the chance to develop their initial opinions, this could be considered undue influence (Tsai et al. 2024). Moreover,

[14] E-mail survey.
[15] This statement is based on a comment from the survey distributed by the authors to the platforms.
[16] E-mail survey.
[17] E-mail survey.

Thomson et al. (2023) examining the democracy platforms DemocracyOS, Consul Democracy, Decidim and D-Cent (all based on open-source technologies) observes that none of them adopts anonymity and full decentralisation technology (such as blockchain) in discussion forums. The authors stress that conventional centralised systems are prone to "single-point of failure problems" and vulnerabilities of data at rest, in transit and storage, which in turn might compromise validity of the content and votes, as well as the trust of the users (Thomson et al. 2023). [18]

Conclusions

Our analysis of selected online deliberation platforms reveals that only some of them actually use AI. The evidence of impact of AI on input and throughput legitimacy also tends to be scarce. When it comes to the inclusiveness of deliberation (input legitimacy), some of the platforms - including *Your Priorities* and *Assembl* - use AI for real-time translation and transcription, which could be facilitating user engagement across linguistic barriers, and hence participation on equal grounds (Landemore 2023). However, we have not found any solid evidence for this type of impact in the literature and the surveyed platforms have not reported any documentation underpinning these claims of improved inclusiveness.

The research on the effect of AI on the throughput dimension of legitimacy yields several key insights. First, the survey among platforms reveals mixed perceptions regarding AI's ability to enhance balanced deliberation, with limited agreement on its efficacy. However, in practice, several platforms are addressing throughput legitimacy with AI features designed to maintain mutual respect, build consensus, synthesis and cluster information in order to support the forming of well-informed preferences. In terms of civility of discussions, platforms like *Your Priorities* and *SODP* are integrating NLP and sentiment analysis to flag and mitigate toxic speech, aiming at fostering mutual respect among users. In terms of consensus-building, platforms such as *Pol.is* and *Ethelo* leverage advanced algorithms to identify areas of agreement and optimise decision-making, employing methods like "Liquid Trust" and neural networks for nuanced participant engagement. Furthermore, platforms like *Assembl* and *Go Vocal* employ machine learning to categorise and visualise discussion topics, facilitating transparent dialogue by organising vast datasets.

The AI support for transparency and accountability of deliberation also seems rather weak, with one of the platforms underlining that one of their current features (AI summarising multiple perspectives into new statements) does rather the opposite. However, several of the surveyed platforms report that they have mechanisms in place to make the AI used on their platform auditable and explainable. Some platforms highlight user-facing transparency features, whereas others focus on the use of open-source models. *YourPriorities* distinguishes itself by providing detailed documentation on the functionality of their AI systems, setting a benchmark for openness in AI operations.

AI does not only bring new opportunities but also new risks for deliberation when it comes to user agency, transparency and auditability of the AI, privacy and anonymity. If the AI is not sufficiently explainable and transparent, and engineers the deliberation process on behalf of users (e.g. by clustering their opinions or by summarising conversations in a certain way), the principle of agency could be threatened and trust of

[18] This paper examines four democracy platforms based on open-source technologies: DemocracyOS, Consul Democracy Decidim and D-Cent. Thomson et al. (2023).

the users undermined. The efficiency of AI might also impact on the levels of creativity and dissent, put aside minority views, and lead to less pluralistic discussions (Moats and Tseng 2023, p. 994).

This chapter has discussed the role and the influence of AI on existing deliberation platforms. However, the democratic legitimacy of the inputs, processes and outcomes also depend on other (non-tech) factors. Fischli and Muldoon (2024) identify a number of such important conditions for the success of experiments in digital democracy, including adaptation to the local (political and organisational) context; institutionalisation to avoid over-reliance on the actions of a single office holder when implementing outcomes, and the concrete passing of decision-making power to citizens; as well as continuous stakeholder engagement (Fischli and Muldoon 2024). The platforms reviewed in this chapter do not seem to challenge or transform the current political institutions or institutional settings. In sum, they are not (*per se*) establishing permanent extra-electoral institutions, nor fundamentally challenging existing power relations, and the deliberative legitimacy concerns are not going to be solved by AI alone.

Acknowledgement

The work of V.Valtenbergs and V.Skolmeistere on this article was funded by the Ministry of Education and Science of Latvia, project "Innovative and inclusive governance for the promotion of social involvement, trust, and communication" (Nr. VPPLETONIKA- 2021/3-0004).

Reference list

Alger, D. Voting by proxy. Public Choice 126, 1–26 (2006). https://doi.org/10.1007/s11127-006-3059-1

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "Nasty Effect:" Online incivility and risk perceptions of emerging technologies. Journal of Computer-Mediated Communication, 19(3), 373–387. https://doi.org/10.1111/jcc4.12009

Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic Talk: How online incivility can undermine perceptions of media. International Journal of Public Opinion Research, 30(1), 156–168. https://doi.org/10.1093/ijpor/edw022

Bjarnason, R., Gambrell, D., Lanthier-Welch. J. (2024). Using Artificial Intelligence to Accelerate Collective Intelligence: Policy Synth and Smarter Crowdsourcing. Paper submitted for ACM Collective Intelligence Conference 2024, Parallel Track Presentation, Boston, MA, USA https://arxiv.org/abs/2407.13960

Buchstein, H. (1996). Bittere Bytes [Bitter Bytes]. Deutsche Zeitschrift Für Philosophie, 44(4), 583–607. https://doi.org/10.1524/dzph.1996.44.4.583

Carstens, J. A., Friess, D. (2023). AI Within Online Discussions: Rational, Civil, Privileged? Ethical Considerations on the Interference of AI in Online Discourse. Minds and Machines (2024) 34:10. https://doi.org/10.1007/s11023-024-09658-0

Citizens. (n.d.). *Your Priorities Feature Overview*. Retrieved October 17, 2024 from https://citizens.is/your-priorities-features-overview/

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Oskar Rydén, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, and Daniel Ziblatt (2024). "V-Dem Codebook v14" Varieties of Democracy (V-Dem) Project.

Chadwick, A. (2011). Explaining the failure of an online citizen engagement initiative: The role of internal institutional variables. Journal of Information Technology & Politics, 8(1), 21-40. https://doi.org/10.1080/19331681.2010.507999.

Davis, M. (2021). The online anti-public sphere. European Journal of Cultural Studies, 24(1), 143–159.https://doi.org/10.1177/1367549420902799

Devenyi, V., Di Giacomo, D., O'Donohoe, C. (2021). *Open Source Software Country Intelligence Report: Argentina*. Wavestone's European Services team, European Union.

Duval, M. (2018, May 28). *vTaiwan : making citizens the key to public debate*. Bluenove. https://www.bluenove.com/en/blog/vtaiwan-making-citizens-the-key-to-public-debate

Ethelo. (2018, January 24). *Use Liquid Trust for Next Level Liquid Democracy*. https://blog.ethelo.org/use-trust-networks-next-level-liquid-democracy

Ethelo. (2021, May 25). *Ethelo Democracy and Artificial Intelligence*. https://blog.ethelo.org/ethelo-democracy-and-artificial-intelligence

Ethelo. (n.d.) *Deep dive into Ethelo's technology*. Retrieved October 15, 2024 from https://ethelo.com/deep-dive/

Freelon, D.G., Kriplean, T., Morgan, J., Bennett, L.W., Borning, A. (2012). Facilitating Diverse Political Engagement with the Living Voters Guide, Journal of Information Technology & Politics, 9:3, 279-297. https://doi.org/10.1080/19331681.2012.665755

Fischli R, Muldoon J. (2024). Empowering Digital Democracy. *Perspectives on Politics*. 2024;22(3):819-835. https://doi.org/10.1017/S1537592724000409.

Friess, D., Ziegele, M., & Heinbach, D. (2021). Collective civic moderation for deliberation? Exploring the links between citizens' organized engagement in comment sections and the deliberative quality

of online discussions. Political Communication, 38(5), 624–646. https://doi.org/10.1080/10584609.2020.1830322

Friess, D., Weinmann, C., & Behrendt, M. (2022). AI and Deliberation. How AI can Support Online Discussions in Deliberative Fashion – a Systematic Review. In 9th European Communication Conference, Aarhus.

Fuller, R. H., Jakovljev, M. (2024). *SDI Digital Democracy Report*. Solonian Democracy Institute.

Fung, A. (2007). "Democratic Theory and Political Science: A Pragmatic Method of Constructive Engagement." American Political Science Review 101 (3): 443–58. https://doi.org/10.1017/S000305540707030X

GitHub. (n.d.). *RightToAskOrg/word_comparison: Comparing questions to look for similar ones*. Retrieved October 15, 2024 from https://github.com/RightToAskOrg/word_comparison

GoVocal. (n.d. a). *Go Vocal: your next digital community engagement platform.* Retrieved October 17, 2020 from https://www.govocal.com/

GoVocal. (n.d. b) *Sensemaking: AI-powered input analysis and summarization*. Retrieved October 16, 2024 from https://www.govocal.com/platform-features/sensemaking

Green-Armytage, J. (2014). Direct voting and proxy voting. Unpublished manuscript. https://web.archive.org/web/20141210071045/http://inside.bard.edu:80/~armytage/proxy.pdf

Jho, W., & Song, K. J. (2015). Institutional and technological determinants of civil e-Participation: Solo or duet?. Government Information Quarterly, 32(4), 488-495. https://doi.org/10.1016/j.giq.2015.09.003.

Landemore, H. (2020). Open Democracy: Reinventing Popular Rule for the Twenty-First Century. Princeton, NJ: Princeton University Press.

Landemore, H. (2023). "Can AI bring deliberative democracy to the masses?" (https://www.law.nyu.edu/sites/default/files/Helen%20Landemore%20Can%20AI%20bring%20deliberative%20democracy%20to%20the%20masses.pdf

Leonard, A. (2020, July 23). *How Taiwan's Unlikely Digital Minister Hacked the Pandemic*. Wired. https://www.wired.com/story/how-taiwans-unlikely-digital-minister-hacked-the-pandemic/

Moats, D., Tseng, Y.-S. (2023) 'Sorting a public? Using quali-quantitative methods to interrogate the role of algorithms in digital democracy platforms', Information, Communication & Society, 27(5), pp. 973–1007. https://doi.org/10.1080/1369118X.2023.2230286

Molteni, M. (30.09.2020). "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/

Papada, E., Altman, D., Angiolillo, F., Gastaldi, L., Köhler, T., Lundstedt, M., Natsika, N., Nord, M., Sato, Y., Wiebrecht, F., Lindberg. S.I. (2023). Defiance in the Face of Autocratization. Democracy Report 2023. University of Gothenburg: Varieties of Democracy Institute (V-Dem Institute).

Rodríguez-Ruiz, J., Mata-Sánchez, J. I., Monroy, R., Loyola-González, O., & López-Cuevas, A. (2020). A one-class classification approach for bot detection on Twitter. Computers & Security, 91, 101715. https://doi.org/10.1016/j.cose.2020.101715

Schmidt, V., Wood, M. (2019). "Conceptualizing Throughput Legitimacy: Procedural Mechanisms of Accountability, Transparency, Inclusiveness and Openness in EU Governance." Public Administration, https://doi.org/10.1111/padm.12615

Stanford Online Deliberation. (n.d). *Online Deliberation Platform*. Retrieved October 17, 2024 from https://stanforddeliberate.org/

Steenbergen, M., Bächtiger, A., Spörndli, M. et al. Measuring Political Deliberation: A Discourse Quality Index. Comp Eur Polit 1, 21–48 (2003). https://doi.org/10.1057/palgrave.cep.6110002

Stephenson P. (2023). Exploring the Throughput Legitimacy of European Union Policy Evaluation: Challenges to Transparency and Inclusiveness in the European Commission's Consultation Procedures and the Implications for Risk Regulation. European Journal of Risk Regulation, 14(2), 351-370. https://doi.org/10.1017/err.2023.33

Stoll, A., Ziegele, M., & Quiring, O. (2020). Detecting impoliteness and incivility in online discussions. Computational Communication Research, 2(1), 109–134. https://doi.org/10.5117/CCR2020.1.005.KATH

Suler, J. (2004). The online disinhibition effect. Cyberpsychology & Behavior, 7(3), 321–326. https://doi.org/10.1089/1094931041291295

Thomson et al. (2023). Direct Democracy System Architecture for Sustainable Community Participation. https://personales.upv.es/thinkmind/dl/conferences/iariacongress/iaria_congress_2023/iaria_congress_2023_1_340_50219.pdf

Tsai, Lily L., Alex Pentland, Alia Braley, Nuole Chen, José Ramón Enríquez, and Anka Reuel. 2024. "Generative AI for Pro-Democracy Platforms." An MIT Exploration of Generative AI, March. https://doi.org/10.21428/e4baedd9.5aaf489a

Wojcieszak, M., Thakur, A., Ferreira Gonçalves, J. F., Casas, A., Menchen-Trevino, E., & Boon, M. (2021). Can AI enhance people's support for online moderation and their openness to dissimilar political views? Journal of Computer-Mediated Communication, 26(4), 223–243. https://doi.org/10.1093/jcmc/zmab006

Van Ransbeeck, W. (2024, September 30). *AI Won't Fix Democratic Participation On Its Own — But It Might Help*. Forbes. https://www.forbes.com/councils/forbestechcouncil/2024/09/30/ai-wont-fix-democratic-participation-on-its-own-but-it-might-help/

Zheng, Y., Schachter, H. L., & Holzer, M. (2014). The impact of government form on eparticipation: A study of New Jersey municipalities. Government Information Quarterly, 31(4), 653-659. https://doi.org/10.1016/j.giq.2014.06.004

Annex 1. Sample: Deliberative democracy platforms

Platforms	**Developers**
Adhocracy+	Liquid Democracy
Ask Parliament	Democracy Developers
Assembl	Bluenove
Citizen Lab	Citizen Lab
Civocracy	Civocracy
Consider.it	Consider.it
Consul Democracy	Consul Democracy Foundation
ConsultVox	Publilegal
CoUrbanize	CoUrbanize
Crowd Law Making	Democracia en Red
Decidim	Decidim Free Software Association
Decision21	Participation21
Dialogue	Delib
Discuto	Impressum
EngagementHQ	Granicus
Ethelo	Ethelo
Fluicity	Fluicity
LiquidFeedback	LiquidFeedback
Loomio	Loomio
Make.org	Make.org
OpenStad	Municipality of Amsterdam
Pol.is	The Computational Democracy Project
Public Consultation	Democracia en Red
Purpoz	Cap Collectif

Stanford Online Deliberation Platform	Deliberative Democracy and Crowdsourced Democracy Teams at Stanford
Wikum	MIT Researchers
Your Priorities	Citizens Foundation

This list represents the deliberative platforms considered in our chapter. Source: https://mit-genai.pubpub.org/pub/mn45hexw/release/1#appendix-2-direct-and-deliberative-democracy-platforms

Annex 2. Summary of the uses of AI in reviewed deliberation platforms based on their self-assessment in the survey

AI functionality	**Description of the functionality as explained in the questionnaire or suggested by the platforms**	**Platforms using this AI functionality**
Text Analysis (NLP)	Automatically identifies key themes and sentiments in citizen feedback, clustering similar ideas and filtering out less relevant inputs. This helps in efficiently summarising and understanding large volumes of data from participants	*Citizen.is/Your Priorities* *Stanford Deliberate* *GoVocal* *Discuto*
Recommendation systems	Provides personalised suggestions for relevant topics and discussions, enhancing user engagement by directing participants to areas of interest based on their previous interactions and preferences	*Citizen.is/*Your Priorities *GoVocal*
Predictive analytics	Utilises data to forecast trends in public opinion and identify potential areas of concern or interest. This helps policymakers anticipate and address emerging issues more proactively	*Citizen.is/Your Priorities*
Automated moderation	Monitors discussions to filter out inappropriate content, ensuring that conversations remain focused and constructive. This reduces the need for extensive human moderation and helps maintain the quality of deliberations	*Citizen.is/Your Priorities* *GoVocal*
Automated reporting	Generates comprehensive summaries and reports of deliberative processes, highlighting key points, decisions, and trends. This makes it easier for both participants and decision-makers to review outcomes and take informed actions	*Citizen.is/Your Priorities* *Stanford Deliberate* *GoVocal* *Discuto*
Translation and linguistic support	Facilitate communication between users, or chatbot feature to help citizens write comments	*Citizen.is/Your Priorities* *Stanford Deliberate* *LiquidFeedback* *GoVocal*
Measuring the quality of deliberation in real-time		*Citizen.is/Your Priorities*
Measure the degree of overlap in the		*Stanford Deliberate* *LiquidFeedback*

content		*Discuto*
Search	Identifying similar questions	*DemocracyDevelopers*
Visualisation of the evolution of the networks	Making it easy to predict convergence of the community around certain views. Building trust networks: voters as vectors (liquid democracy)	*Discuto*
Clustering	Clustering of users with similar opinions	*Discuto*