

## RESEARCH ARTICLE

# Integrating Artificial Intelligence Into Citizens' Assemblies: Benefits, Concerns and Future Pathways

Sammy McKinney

Interest in how Artificial Intelligence (AI) could be used within citizens' assemblies (CAs) is emerging amongst scholars and practitioners alike. In this paper, I make four contributions at the intersection of these burgeoning fields. First, I propose an analytical framework to guide evaluations of the benefits and limitations of AI applications in CAs. Second, I map out eleven ways that AI, especially large language models (LLMs), could be used across a CAs full lifecycle. This introduces novel ideas for AI integration into the literature and synthesises existing proposals to provide the most detailed analytical breakdown of AI applications in CAs to date. Third, drawing on relevant literature, four key informant interviews, and the Global Assembly on the Ecological and Climate crisis as a case study, I apply my analytical framework to assess the desirability of each application. This provides insight into how AI could be deployed to address existing challenges facing CAs today as well as the concerns that arise with AI integration. Fourth, bringing my analyses together, I argue that AI integration into CAs brings the potential to enhance their democratic quality and institutional capacity, but realising this requires the deliberative community to proceed cautiously, effectively navigate challenging trade-offs, and mitigate important concerns that arise with AI integration. Ultimately, this paper provides a foundation that can guide future research concerning AI integration into CAs and other forms of democratic innovation.

**Keywords:** Artificial Intelligence; Large Language Models; Citizens' Assemblies; Deliberative Mini-Publics and Citizen Participation

## Introduction

Democratic innovations are novel processes that strengthen the role of lay citizens in governance processes through providing 'opportunities for deliberation, participation and influence' (Elstub & Escobar 2019: 11). Citizens' Assemblies (CAs) – (quasi)randomly selected groups of citizens brought together to learn about, deliberate on, and reach recommendations on political issues – are a prominent kind of democratic innovation that shows promise in grappling with complex governance questions whilst promoting democratic renewal. Resultingly, they are increasingly in the public eye, and subject to ongoing academic research.

Artificial Intelligence (AI) is also increasingly in the public eye, particularly following the release of OpenAI's large language model (LLM) ChatGPT. Artificial Intelligence has a complex relationship with democracy: whilst it can threaten democratic legitimacy (Danaher 2016) and erode the citizenry's access to reliable knowledge (Coeckelbergh 2022), it can also improve the policy cycle (Craglia, Hradec & Trousaard 2020) and enable mass deliberation (Landemore 2022).

Widespread interest in the role that AI could play in CAs is emerging amongst scholars (e.g., Landemore 2022)

and practitioners (e.g., Devine 2023) alike. This paper responds to this increasing interest through making four contributions at the intersections of these burgeoning fields.

First, I propose an analytical framework to guide evaluations of the benefits and limitations of AI applications in CAs. Second, I map out eleven ways that AI could be used across a CAs full lifecycle. This introduces novel ideas for integrating AI into CAs and synthesises existing proposals to provide the most detailed analytical breakdown of AI applications in CAs to date. These applications are summarised below in **Table 1**.

Third, drawing on relevant literature, four key informant interviews, and the Global Assembly on the Ecological and Climate crisis as a case study, I apply my analytical framework to assess the benefits and limitations of each application. Fourth, I bring my critical analysis together to argue that AI integration into CAs brings the potential to enhance their democratic quality and institutional capacity. Realising this, however, requires the deliberative community to effectively navigate challenging trade-offs and mitigate important concerns that arise with AI integration. A key trade-off is deploying AI in ways that can improve the institutional capacity of CAs without undermining their democratic quality; and important concerns may include bias, 'hallucinations', limited

**Table 1:** Overview of Eleven Possible AI Applications in CAs.

Recruitment and Planning	Learning	Deliberation/Decision-Making	Follow-up
<ul style="list-style-type: none"> <li>Clustering input to support agenda setting</li> </ul>	<ul style="list-style-type: none"> <li>Simplifying and summarising learning materials</li> <li>Assisting with question generation</li> <li>Acting as question-and-answer systems</li> </ul>	<ul style="list-style-type: none"> <li>Facilitation</li> <li>Measuring deliberative quality</li> <li>Playing devil's advocate</li> <li>Aggregating across deliberations</li> <li>Generating consensus statements</li> </ul>	<ul style="list-style-type: none"> <li>Communicating output with the public</li> </ul>
		<ul style="list-style-type: none"> <li>Translation</li> </ul>	

transparency, and excessive machine influence. Proceeding with caution, deploying AI in hybrid formats, and taking measures to mitigate concerns are essential steps for meeting these challenges and promoting a positive future for tech-enhanced CAs.

This paper proceeds in seven steps. First, I outline my methodology. Second, I provide an overview of relevant literature on CAs and AI. Third, I propose an analytical framework for assessing the desirability of integrating AI applications into CAs. Fourth, I introduce a call to caution concerning AI integration into CAs by exploring the concerns of AI bias, hallucination, and excessive machine influence. Fifth, I apply my analytical framework to eleven different AI applications spanning the entire lifecycle of CAs, providing an analytical breakdown of how AI, especially LLMs, can be used in CAs and an analysis of the benefits and limitations of each application. Sixth, I bring my analyses together to reflect on the overarching democratic and institutional implications of AI integration. Seven, I consider how my findings generalise beyond CAs to other kinds of democratic innovation and identify areas for future research.

### Methodology

There were three key dimensions to my methodological approach. First, I conducted 'key informant' interviews: detailed conversations with individuals possessing expert knowledge on a specific topic (Taylor & Blake 2015: 153). My four key informants were Kelly McBride, Yago Bermejo, Claire Mellier, and Johnny Stormonth-Darling. These practitioners have collective experience designing and implementing CAs at national, transnational, and global levels (appendix A provides an overview of their experience).

Key informant interviews were selected because they can provide high quality and novel insight on a topic in a short timeframe (Taylor & Blake 2015). Specifically, my key informants' experience in designing and delivering CAs means they are well situated to identify the extent to which AI applications can address certain real-world challenges facing CAs today.

Second, the Global Assembly on the Climate and Ecological crisis (henceforth Global Assembly) is used as a case study throughout. This was selected for three reasons: global deliberation is a 'frontier' in deliberative democracy research (Dryzek 2012); a recent evaluation of the Global

Assembly provided rich insights to build on (Curato et al. 2023); and three of my key informants were involved in the Global Assembly's design and implementation.

Third, combining this case study, qualitative data from my key informant interviews, and other relevant literature, I conducted applied normative theorising surrounding how AI could be integrated into CAs, and the implications of doing so. This approach follows what Chambers (2022) conceptualises as 'critical applied theory' and 'applied constructive theory': I critically evaluated existing approaches to CAs, and constructively proposed ways of improving them, in light of normative theory.

### Citizens' Assemblies and Artificial Intelligence

Citizens' Assemblies are a type of deliberative mini-public. Whilst there is not one accepted definition of deliberative mini-publics, Ryan and Smith (2014: 20) propose that they must be constituted by an 'inclusive and representative sub-group of an affected population' and engage in 'structured deliberation enabled by independent facilitation'. The first condition concerns who participates: deliberative mini-publics are composed of diverse citizens representing a cross-section of those affected by the issue at stake. Recruitment strategies for achieving this vary, but it often occurs through combining random selection with stratification according to socio-demographic and attitudinal criteria (Gašiorowska 2023: 5). The second condition concerns how participants engage with the topic of interest: through facilitated deliberations. Deliberation is a specific form of communication that 'involves weighing and reflecting on preferences, values, and interests regarding matters of common concern' (Bächtiger et al. 2018: 2). It places emphasis on listening openly and sharing public reasons (reasons that others can plausibly accept) (MacKenzie 2018: 253).

CAs tend to include 99–150 participants, last several weekends, and create a recommendation on the topic at issue (Setälä & Smith 2018: 301). Generally, CAs have five key phases (Escobar & Elstub 2017: 4–5), which I have broken down in **Table 2**. These phases make up what I refer to as the 'lifecycle' of CAs.<sup>1</sup>

CAs are grounded in the theory of deliberative democracy. Deliberative democracy is rooted in the ideal of equal citizens respectfully reasoning together on issues of common concern, and through that reasoning formulating policies that affect their lives (Bächtiger

**Table 2:** Five Key Phases of Citizens' Assemblies.

Phase	What happens?
Recruitment and planning	The organisers of the assembly plan the process and a representative cross-section of the affected population are recruited through a process of stratified random selection.
Learning	Participants learn about the topic through engaging with learning materials and relevant speakers, such as experts, stakeholders, and activists.
Deliberation	Supported by facilitators, participants deliberate on the topic by listening to others and exchanging public reasons.
Decision-making	Through group consensus or voting, the assembly formalises their recommendations.
Follow-up	The outcomes of the CA are shared with relevant actors, such as legislatures and the wider public.

et al. 2018: 2). CAs seek to create conditions that are conducive to realising this ideal and, as such, are one of the most active research areas within deliberative theory and practice.

Like CAs, AI also lacks a shared definition (Wang 2019). Despite this, it has been suggested that definitions often overlap in viewing AI as 'agents (programs running on computer systems) that can learn, adapt, and deploy themselves successfully in dynamic and uncertain environments' (Miaillhe & Hodes 2017: 3). Machine learning algorithms, for example, use statistical methods to learn correlations in datasets that enable them to predict real-world outcomes.

An important development in AI has concerned large language models (LLMs), such as ChatGPT. LLMs are deep neural networks that use huge corpuses of training data to generate novel textual output through predicting the likelihood of the next token, typically a word. LLMs can serve various functions, such as summarisation (Zhang et al. 2023), translation (Lyu et al. 2023), and identifying agreement amongst diverse perspectives (Bakker et al. 2022). Notably, however, there are several concerns surrounding AI systems, including LLMs, such as bias (e.g., Bender et al. 2021), privacy (e.g., Tucker 2018), and transparency (e.g., von Eschenbach 2021).

There is a growing body of research at the intersection of AI and democratic innovations. Stanford's Deliberative Democracy lab is experimenting with online deliberation that uses AI facilitation (Fishkin et al. 2018). It has been found that artificial facilitators can enhance interactions in online deliberation (Wyss & Beste 2017), but that they lack the ability to justify their behaviour and may be problematically imposed by experts (Alnemr 2020). Small et al. (2023) explore the way LLMs could be integrated into Polis, such as to identify topics, summarise content and generate consensus statements. Polis is an open-source crowdsourcing platform that takes large amounts of public input on a topic and, through machine learning algorithms, clusters viewpoints. The computational democracy project (2023) has surveyed how Polis could be integrated into CAs, proposing that it could be used to democratise agenda-setting and to distribute deliberations to the wider public. Furthermore, Landmore (2022) proposes different ways AI could be integrated into deliberative mini-publics, including for translation, facilitation, fact-checking, and

clustering viewpoints. She argues that this could support deliberation at mass scale.

Building on this research, this paper aims to advance our understanding of AI integration into democratic innovations, especially CAs. It does so by: (a) proposing an analytical framework for evaluating AI integration into CAs and (b) providing a detailed map and analysis of how AI, especially LLMs, could be deployed within CAs throughout their lifecycle. I explore how my findings generalise beyond CAs to other kinds of democratic innovations in the penultimate section.

### An Analytical Framework

Smith (2009) provides an analytical framework for assessing democratic innovations. This framework identifies democratic and institutional 'goods': virtues that are relevant to the democratic quality and institutional effectiveness of CAs. I propose that assessing whether AI applications promote rather than hinder these democratic and institutional goods is key to a robust critical analysis of the desirability of integrating AI into CAs. Democratic goods are important because CAs are democratic processes, meaning the effects AI integration has on their democratic quality is normatively significant. Institutional goods are important because if we focus exclusively on whether AI integration promotes democratic ideals, then we fail to consider what AI integration means for the real-world institutionalisation and effectiveness of CAs.

Adapting Smith's (2009) framework, I propose four democratic goods and two institutional goods for consideration, broken down in **Table 3**:

This framework provides a guide for assessing the desirability of integrating AI into CAs according to the two crucial dimensions of democratic quality and institutional capacity.<sup>2</sup> Of course, AI applications do not need to promote all these goods to be deemed desirable. Rather, this framework provides a way to identify the benefits and limitations of a given AI application, which is instructive for judging their desirability.

### A Call to Caution: Bias, Hallucinations, and Excessive Machine Influence

Before mapping and analysing specific AI applications in CAs, I would like to introduce a call to caution. I display the importance of proceeding cautiously through exploring three important concerns surrounding the integration of

**Table 3: Democratic and Institutional Goods: An Analytical Framework for Assessing AI applications in CAs.**

<b>Democratic Goods</b>	
Inclusiveness	Does the AI application promote the equal inclusion of those whose interests are affected by the topic at hand?
Popular Control	Does the AI application promote the ability of participants and the wider public to have policy impact and exercise control over the decision-making process?
Considered Judgement	Does the AI application support participants in reaching 'thoughtful and reflective judgements' (Smith 2009: 24)?
Transparency	Does the AI application promote the ability of participants and the wider public to understand the process, such as how citizen perspectives are aggregated?
<b>Institutional Goods</b>	
Efficiency	Does the AI application reduce the economic and social costs of CAs?
Scalability	Does the AI application promote higher levels of citizen participation and the scalability of CAs to transnational levels?

AI into CAs: bias, 'hallucinations', and excessive machine influence. These concerns, as I show with reference to my analytical framework, affect the democratic quality of CAs and are therefore crucial to address.

If biases are present in an AI system's training data, they are sustained in the model's output (e.g., Bender et al. 2021). Artificial intelligence bias can refer to AI systems leading to disparate treatment between individuals or groups without a reasonable justification for such a disparity (Kordzadeh & Ghasemaghaei 2022: 388). This has been seen in relation to, for example, race (e.g., Benjamin 2019) and gender (e.g., Leavy 2018). Alongside disparities in treatment, AI bias can also take the form of AI systems having certain political leanings, such as left-wing biases (Hartmann et al. 2023). To apply my framework, AI creating disparities in treatment between groups in a process would undermine inclusion, and AI favouring certain worldviews would lead to an unbalanced provision of knowledge, undermining considered judgement.

A second issue is so-called hallucinations in AI systems. Hallucinations refer to AI systems generating incorrect information through 'making things up when they don't know something' (Small et al. 2023: 6). The capacity of AI systems to hallucinate has been identified in various applications, such as in translation (Guerreiro et al. 2023). Such information errors would undermine the ability of participants to reach considered judgement.

To lessen concerns of bias and hallucination, the designers of AI systems can 'fine-tune' AI models to improve their performance, such as on specialised datasets (e.g., de Vassimon Manela et al. 2021). Furthermore, there needs to be relevant human oversight, or a 'human-in-the loop', to review, verify, and add nuance to AI generated content. This human-in-the-loop approach was highlighted repeatedly by the interviewees of Bermejo and Stormonth-Darling as being essential for addressing biases and hallucinations when deploying AI in CAs, and research has also stressed its importance (e.g., Fayyad 2023). As such, maintaining human oversight and conducting model fine-tuning are important steps for mitigating concerns of bias and hallucination.

A third issue with AI integration is ensuring that AI is not deployed in ways that gives too much influence to AI systems

over human deliberations. As Small et al. (2023: 16) ask: 'How much machine influence is acceptable in a process which ultimately aims to surface human opinions?' Underpinning this concern is the issue of popular control: AI systems should not be deployed in ways that undermine the ability of citizens themselves to shape and ultimately determine the outcomes of CAs. Instead, AI integration should be used to empower and support human participation, rather than replace it. As such, there should always be opportunities for humans to verify, contextualise and critically engage with AI generated material, thereby ensuring human deliberators are responsible for the CA's output.

Scoping out these three concerns highlights the importance of proceeding cautiously with AI integration into CAs: if sufficient safeguards are not put in place, AI integration could significantly damage public perceptions of CAs and undermine their democratic quality. The deliberative community must be attentive to concerns that arise with AI integration, aware of steps that can be taken to mitigate them and carefully weigh up whether the risks and shortcomings are worth bearing. In what follows, I explore specific AI applications, noting when concerns like these arise and highlighting how they may be alleviated.

### **Mapping and Analysing AI applications in CAs**

I will now map out and analyse eleven AI applications across the full lifecycle of CAs. I focus predominantly on applications that deploy LLMs due to their recent advancement and ability to execute a wide range of tasks.<sup>3</sup> I take state-of-the-art LLMs like OpenAI's GPT-4 or Anthropic's Claude as the object of analysis. I start by discussing AI translation, since this application can be used across the entire lifecycle of CAs. I then consider LLM applications across four key phases of CAs: planning and recruitment, learning, deliberation, and follow-up.

#### **Translation**

Translation is essential to many CAs, particularly those at the transnational level, to allow participants to converse across linguistic divides. In the Global Assembly a consecutive translation approach was used where speakers would pause to give translators time to translate.

This approach to translation in the Global Assembly faced several issues, such as insufficient capacity to translate certain learning materials (Global Assembly Team 2022: 83); unequal quality of translation across participants (Curato et al. 2023: 102–103); and translators not having the time to translate nuance, stories, and relevant context (Ibid). For these reasons, Curato et al. (2023: 111) suggest that the use of digital tools for live-time translation should be explored for future global assemblies.

AI translation can offer four main benefits to CAs. First, it can operate in live time, unlike the consecutive translation of the Global Assembly, giving participants more time to deliberate and reach considered judgement. Second, it is more practical and less expensive than hiring 'an army of translators', the cost of which can be 'rather prohibitive' (Landemore 2022: 29). As such, AI translation can increase the efficiency of CAs by lowering the cost of translation. Third, due to the importance of translation to transnational and global level CAs, AI translation could increase the scalability of CAs. Fourth, AI translation could enable a transition from English being the exchange language of global deliberative processes. This would increase inclusiveness: using English as the Global Assembly's exchange language created inequalities because English speakers had easier access to information (Global Assembly Team 2022: 174).

However, there are two important concerns surrounding AI translation. First, whilst AI translation is good for high-resource languages – languages with rich data resources available, like English – they are not as effective at translating low-resource languages (Hendy et al. 2023). The vast majority of languages in the world are low or extremely low resource (Huang et al. 2023: 2). The discrepancy in translation quality between high-resource and low-resource languages is the result of some languages having higher quality training data for AI systems than others (Nicholas & Bhatia 2023: 6). This discrepancy would exacerbate inequalities within CAs: speakers of the world's high-resource languages will receive higher quality translation compared to speakers of low-resource languages. So, whilst AI systems may be deployed to address Anglocentrism in CAs, their deployment could exacerbate it.

The second shortcoming concerns political translation (Doerr 2018). Political translation occurs when third party actors attempt to address inequalities in deliberative settings by advocating for disadvantaged groups. During multilingual deliberations, political translation often entails translating between participants in a way that contextualises background inequalities, rather than providing a literal translation of speech. For Curato et al. (2023: 111), such political translation is important for empowering the disadvantaged, therefore promoting inclusiveness. However, AI systems cannot be expected to perform effective political translation (Landemore 2022: 29). As such, AI translation would not promote more inclusive deliberations as human-enabled political translation.

Given these shortcomings, caution is required with AI translation and a nuanced hybrid approach to AI translation is required to promote its benefits whilst

mitigating its risks. It will be particularly important to use human translation when low-resource language or political translation are required during multilingual deliberations. However, AI translation could still be useful throughout the CA lifecycle to translate materials, such as learning documents, invitation letters, and social media posts, which could then be verified and nuanced by humans. This hybrid approach would enable AI translation to promote the efficiency and scalability of CAs without undermining the inclusiveness of deliberations.

### **Planning and recruitment**

#### **Democratising agenda-setting**

During planning, the practitioners responsible for designing and delivering CAs create an overview of what the CA is going to look like. One important aspect of planning concerns the role that the wider public should play a part in agenda-setting. It is well established in social science that agenda-setting influences outcomes. The framing of the Global Assembly, for example, potentially skewed the agenda towards thinking about climate change mitigation, opposed to adaptation, which may have benefitted the better off (Curato et al. 2023: 66). For reasons like this, scholars have called for increased democratisation over the agenda-setting of CAs (Ibid).

The Computational Democracy Project (2023) proposes that Polis could be used to support democratised agenda-setting in two ways. First, the public can give input on what issues they would like a CA to address. Second, once a topic is decided, the public can help frame the issue by giving their perspective on the topic, allowing for the identification of key agreements, disagreements, and perspectives across the citizenry. For both applications, LLMs could cluster citizens' contributions into a digestible format, enabling mass input to be effectively analysed. As such, agenda-setting power is distributed to the citizenry, or 'maxi-public', increasing popular control and inclusion.

However, an important concern with this application is that political participation across a range of processes is 'strongly positively correlated with income, wealth and education' (Smith 2009: 14). So, channelling AI to democratise agenda-setting may result in outcomes that favour privileged groups, without truly deepening popular control and inclusion. This highlights the importance of incentivising wide participation in AI-supported crowdsourcing efforts, such as through effective communication (Aitamurto & Landemore 2015: 9).

### **Learning**

#### **Presenting information in accessible formats**

It is essential for CAs to not just ensure that diverse sources of information are presented to participants, but also that this information is presented in diverse formats (Curato et al. 2021: 80). This is because it helps promote more effective learning and accommodates neurodivergence, promoting considered judgement and inclusion. This is why, for example, deliberative processes have used simplified language to support participants with learning disabilities (Scottish Learning Disabilities Observatory 2021).

Here are two ways LLMs could be useful for presenting information in accessible formats:

- Simplification. LLMs can take complex human language and simplify it. Feng et al. (2023) found that LLMs were as effective as human annotators in simplifying sentences.
- Summarisation. LLMs are increasingly effective at performing summarisation tasks. For example, they can be as effective as humans at summarising news (Zhang 2023). This suggests LLMs could be used to provide summaries in the learning phase, for example by summarising the key points from an information booklet or speaker presentation.

Using LLMs to present information in accessible formats was met with positive responses by the interviewees. Mellier mentioned that, as it stands, the knowledge provision within most CAs is problematic because it does not cater to diverse learning styles. Bermejo, furthermore, highlighted that experts often struggle to present information in ways that are digestible to lay citizens. Using LLMs for simplification and summarisation could be helpful in addressing this.

However, biases and hallucinations are an issue for LLM summarisations, meaning that human oversight will be essential to mitigate these concerns (Small et al. 2023: 13). With this necessary precaution, using LLMs to simplify and summarise the key points presented by experts could be valuable to promoting inclusiveness, through presenting information in accessible ways, and promoting considered judgement, through supporting more effective learning.

### **Assisting with question generation**

It is important that participants do not just learn from speakers, but that they can also challenge and critique them. However, empowering participants to do this can be difficult. In the Global Assembly, for example, an unequal power dynamic between participants and speakers arose, with many participants feeling like they were pupils in an environment akin to a school (Curato et al. 2023: 70). To address this, LLMs could assist participants in generating ideas for how to challenge speakers. To exemplify:

**You**

In the Global Assembly, a speaker has just defended the position of ecological modernisation. Can you suggest one question I could pose to the speaker to challenge this position? Make sure the question is written in simple language.

**ChatGPT**

Certainly! Here's a straightforward question you could ask to challenge the position of ecological modernization:

"How can we ensure that ecological modernization doesn't just lead to 'greenwashing,' where companies appear environmentally friendly without truly changing their harmful practices?"

It is important that LLM generated questions are verified by assembly staff and that they only play an assistive role: they should provide content for participants to critically engage with and build on, thereby empowering, rather than replacing, participants when asking questions. Taking

these precautions, using LLMs to assist with question generation could promote considered judgment in CAs by deepening interactions with speakers.

### **'On tap' information sources**

A third application to consider is using LLMs as 'on tap' (Sprain et al. 2014) sources of information for participants. This means that LLMs could be available at any time to support participant learning through acting as question-and-answer systems when relevant experts are not present. McBride mentioned that it is difficult for CAs to get responses to questions in a 'timely and effective manner'. This was a notable problem for the Global Assembly: participants could not robustly scrutinize the evidence presented because witnesses were not always there to directly engage in dialogue (Curato et al. 2023: 76). So, using LLMs as 'on tap' sources of information, Mellier suggested, could have been helpful to address this.

However, using AI as an 'on tap' information source is an especially high-risk application. This is because it is particularly susceptible to undermining considered judgement: AI bias could undermine a balanced provision of knowledge and hallucinations could lead to information errors. Of course, 'experts have biases and blind spots', too (Curato et al. 2021: 74). However, as Stormonth-Darling and Bermejo suggested, we could not trust LLMs to play key roles in knowledge provision without model fine-tuning and relevant human oversight to verify and nuance AI generated content.

As an additional precaution beyond fine-tuning and human oversight, critical engagement sessions could be integrated into the CAs schedule (Carson 2017). These would inform participants on how to critically engage with AI generated material, such as familiarising them with the issues of AI bias and hallucination. Taking these steps, using LLMs as 'on tap' information sources could promote the institutional capacity of CAs whilst maintaining considered judgement, but caution is essential.

### **Deliberation**

#### **AI facilitation**

The state-of-the-art for AI facilitation in deliberative contexts is an artificial facilitator that has been integrated into Stanford's Online Deliberation Platform (Fishkin et al. 2018). This artificial facilitator can perform several functions, such as promoting equitable speaking time, detecting toxic behaviour, and identifying stalls in conversations.

AI facilitation may be desirable because human facilitators are often biased, and AI facilitation could provide a more consistent and impartial experience (Landmore 2022: 28). However, the key virtue of AI facilitation is arguably that it can scale deliberative processes. For example, Stanford's artificial facilitator enabled 2,069 small groups to deliberate (Chang et al. 2023: 7). Without the platform, the process would have required over 500 human moderators operating across 19 languages, which would be logistically challenging and expensive (Chang et al. 2023: 8). This suggests that AI facilitation may be essential for deliberation at scale.

However, facilitation is more akin to a craft than a rigid discipline (Escobar 2019: 182). Facilitators operate in dynamic environments that require them to be attuned to subtleties in communication, such as body language, tone of voice, and cultural sensitivities. They must be trusted as 'honest brokers' (Escobar 2019: 184) and they play an essential role in ensuring all individuals feel comfortable contributing to deliberations. Facilitators can promote this, for example, through deploying humour (Hewer et al. 2019). Furthermore, facilitators should be sensitive to inequalities in the group and can play an important role in their mitigation (Asenbaum 2016). In transnational deliberations, for example, facilitators should be attentive to 'the multiplicity of knowledge, forms of communication, inequalities, deliberative traditions, and individual/communal types of political engagement present in a transnational setting' (Curato et al. 2023: 80). Arguably, this ability of facilitators to be sensitive to nuances in communication, to foster trust and group cohesion, and to mitigate inequalities is an importantly human craft that AI systems cannot mirror. AI facilitation may then undermine inclusion through not being sensitive to inequalities and considered judgement through struggling to promote robust deliberations. Resultingly, as McBride and Bermejo stressed, AI facilitation will not be as effective as human facilitation.

Given these shortcomings, whilst AI facilitation may enhance scalability, it could significantly undermine the democratic quality of CAs. To navigate this tension, there are important hybrid models to consider. First, AI could facilitate some groups in CAs, whilst humans facilitate others, and the groups could be regularly rotated between them. This could improve the institutional capacity of CAs without incurring as high a cost to democratic quality as full AI facilitation would.

Second, human facilitators could be supported by an AI system. Whilst this would not bring the scalability benefits that AI facilitation would bring, it carries the potential to increase the quality of deliberation and promote considered judgement. Mellier, for example, suggested that AI could spot patterns in deliberations that a human may miss. Here, live-time deliberative analysis and other proposed applications discussed below could be used to prompt human facilitators, thereby providing a hybrid form of facilitation.

### **Measuring deliberative quality**

As Landemore (2022: 31) proposes, AI systems could be used to measure the quality of deliberation in live-time. Mellier suggested that such an application would have been 'invaluable' for the Global Assembly. Live-time deliberative analysis could be achieved through training an AI system on existing measures of deliberative quality. For example, LLMs could monitor whether contributions promote or hinder deliberative quality following a Deliberative Transformative Moments approach (Jaramillo & Steiner 2014). This real-time feedback would be available to facilitators who could use it to deepen deliberations and address deliberative shortcomings

as they arise, promoting inclusion and considered judgment.

### **Devil's advocate**

LLMs could sharpen the quality of a CA's output through playing devil's advocate. As Mellier mentioned, it is possible within deliberations that 'groupthink' and uncritical consensus can emerge. To address this, LLMs could produce objections that are integrated into the assembly to test the emerging consensus and thereby promote considered judgement. This role of devil's advocate is sometimes played by humans to good effect (Roberts & Escobar 2015: 86).

This role of devil's advocate entails risk, such as potentially pushing specific perspectives on participants, and therefore requires caution. An important principle to guide this application, suggested by McBride, is that the AI generated content should be framed as open-ended questions to ensure that it does not push participants in certain directions. To exemplify how this application could function, consider the following:

**You**

There is an emerging consensus in the global assembly that ecocide should be enforced globally. Please provide one short, easily understandable objection to this view. Phrase it as an open ended question and ensure the language is as simple as possible.

**ChatGPT**

"Is it fair to expect every country, regardless of its circumstances, to enforce the same rules on ecocide globally, without considering their individual needs and capabilities?"

🗑️ 🔄 🗣️

### **Aggregating across deliberations**

Currently within CAs, assembly staff summarise deliberations and synthesise these summaries into one aggregated output reflecting the perspective of the entire assembly. However, this approach encounters an issue of scale: it is very challenging for humans to synthesise vast amounts of input. For example, the Global Assembly coupled with local level climate assemblies to open their process to wider public input. However, they could not develop a mechanism to 'aggregate the qualitative findings from the many Community Assemblies' (Global Assembly Team 2022: 203). This highlights the difficulty that CAs face in aggregating insights across many deliberation groups.

Here, LLMs could be useful: unlike humans, they can take large amounts of qualitative data to cluster viewpoints in a way that identifies key perspective, agreements, and disagreements. Polis is exploring using LLMs to perform this function in synthesising mass input (Small et al. 2023). This aggregation function is particularly important for promoting scalability because it could enable more people to participate in deliberation whilst still ensuring meaningful synthesis of their perspectives.

However, this application must be approached critically: the act of synthesising input is essential to deliberative processes because it is 'literally the mechanism by which meaning is made' (Small et al. 2023: 11). One important concern is AI systems generating errors or introducing

biases through the aggregation process. An additional issue, raised by McBride, surrounds transparency within an AI driven aggregation process. When humans perform these roles, they can display their workings on whiteboards or spreadsheets in front of participants, allowing participants to oversee the aggregation process. However, the way LLMs perform this role would be less intelligible to citizens, undermining transparency.

To mitigate these concerns, AI aggregation systems must be fine-tuned and retain human oversight. Human oversight can be achieved through always presenting the AI generated aggregations back to participants for them to comment on and correct (Small et al. 2023: 13). These precautions would increase the accuracy and transparency of AI aggregation.

### Generating Consensus Statements

LLMs could be used to propose high-potential ideas in CAs (Landmore 2022: 32), such as consensus statements. LLMs have already shown the ability to propose statements that can garner consensus in a group containing diverse viewpoints (Bakker et al. 2022). This means that LLMs could support participants to find common ground and resolve conflicts more quickly. This is important because limited time is a recurring issue for CAs. At the Global Assembly, for example, participants did not have the time to deeply explore the implications of some proposals (Curato et al. 2023: 78). Plausibly, using LLMs to generate consensus statement could speed up the deliberation process whilst identifying high-potential ideas for consideration, therefore increasing efficiency, and considered judgement.

However, problematic content generation is an important concern. As Landmore (2022: 32–33) asks: 'what if the AI... helps spread populist wildfires rather than reasonable consensus?' To guard against this, human oversight from assembly staff will be important: they could check the statements prior to integrating them into the assembly to prevent harmful ideas spreading.

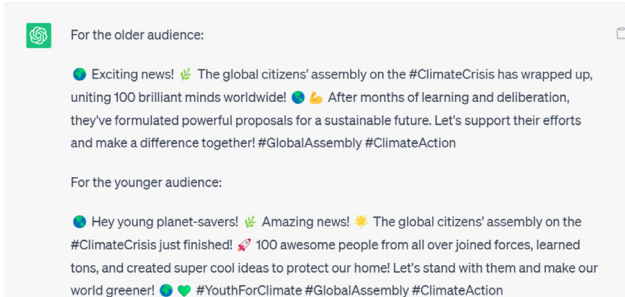
A second important concern was voiced by McBride: 'at what point does the AI become a participant?'. This highlights the important concern of AI systems having too much influence in human deliberations and thereby undermining popular control. To address this, any machine generated idea must be subject to critical deliberation amongst participants, rather than being uncritically accepted, to ensure the output is determined by the citizens themselves.

### Follow-Up

#### Supporting effective communication

LLMs could assist in supporting effective communication with the wider public during the follow-up phase. As suggested by Mellier, LLMs could enable a move beyond a one-size-fits-all approach to communication to one that is tailored to different groups. For example, LLMs could generate different versions of text to appeal to different audiences, such as based on age or political leanings. To exemplify:

SA We have just concluded a global citizens' assembly on the climate crisis that brought together 100 people from all over the world. They learnt, deliberated and formulated several proposals for how to effectively and fairly address the climate crisis over many months. Please create two short tweets to share this news, one for an older audience and one for a younger audience.



Using LLMs to support effective communication with the maxi-public is important for sharing information about CAs efficiently at scale and for promoting their transparency and impact. However, due to biases and hallucinations, AI generated content would need to be verified and refined by humans.

### Bringing it all Together: The Democratic and Institutional Implications of AI Integration

In deploying my framework, the eleven AI applications discussed have been analysed according to their impact on the democratic and institutional quality of CAs. So, what does this critical analysis tell us about the democratic and institutional implications of AI integration into CAs?

We have seen that whilst there are ways AI applications can enhance the democratic goods of inclusiveness, considered judgement, popular control, and transparency in CAs, there are also ways they can undermine these goods. This highlights that if AI integration is to enhance the democratic quality of CAs, steps must be taken to mitigate the ways AI applications can undermine democratic goods. To reflect this, I highlighted necessary precautions to be taken when such concerns arise. For example, we should fine-tune AI models and keep relevant humans-in-the-loop to mitigate bias and hallucination undermining considered judgement.

Furthermore, AI applications, notably translation, facilitation, and aggregation, could significantly enhance the institutional capacity of CAs. On this, my analysis aligns with Landmore (2022), who argues that these applications are promising for scaling deliberation. However, it was also found that these applications bring important concerns for democratic quality. For example, fully automated AI facilitation is unlikely to be as effective as a skilled human facilitator in performing the craft of facilitation, raising doubts about the capacity of AI facilitation to promote inclusiveness and considered judgement within deliberations.

This means that if AI integration is to serve robust democratic *and* institutional ends, the deliberative community will have to navigate these challenging trade-offs. To do so, hybrid possibilities for AI translation and AI facilitation should be explored whereby the institutional capacity of CAs could be advanced without incurring a high cost to democratic quality. For example, human translation can be used for low-resource and political translation



during deliberations, and AI translation with human oversight can be used at other points in the CA lifecycle. This hybrid approach could promote institutional capacity whilst maintaining democratic quality. This suggests that, if the deliberative community engages with these challenging trade-offs and takes steps to mitigate concerns that arise with AI integration, AI integration could enhance the democratic and institutional potential of CAs.

### Beyond CAs

Whilst I have focused on CAs due to their increasing prominence, there is some direct transferability in my findings between CAs and other forms of democratic innovation, like participatory budgeting and digital engagement processes. For example, LLMs could plausibly be used to support translation, summarisation, and the identification of consensus in all of these processes. Furthermore, the concerns I have explored, including bias, hallucinations, and excessive machine influence, will also apply to AI deployment in other forms of democratic innovation. This means that if AI integration is to promote the democratic and institutional capacity of other modes of democratic innovation we must also proceed cautiously and take steps to mitigate these concerns.

However, other forms of democratic innovations will also benefit from AI applications specific to their needs. For example, digital engagement processes could use LLMs for comment routing (Small et al. 2023). This is the process of determining which comments should be displayed to participants to yield greater insights and engagement. Participatory budgeting could explore, for example, using LLMs to support analyses of the trade-offs in budget proposals. So, whilst there is some transferability between the discussed applications in CAs to other democratic innovations, different processes have their own needs, and this will often warrant adapted or specialised AI applications. As such, future research will benefit from exploring and critically analysing the AI applications most appropriate to other forms of democratic innovations.

### Conclusion

This paper aimed to advance our understanding of how AI can be integrated into CAs and the democratic and institutional implications of doing so. It has provided an analytical framework for assessing AI integration into CAs and offered a breakdown and analysis of eleven different AI applications. It was argued that AI integration into CAs can enhance their democratic quality and institutional capacity, but that there are important trade-offs to navigate and concerns to mitigate if AI integration is to serve robust democratic and institutional ends. As such, whilst it will be exciting to see interested actors experiment with AI integration in CAs and other forms of democratic innovation, we must remember to proceed with caution.

One unexplored issue of high significance is the political economy underpinning the AI tools integrated into democratic innovations. Whether these AI tools are provided through public or private ownership will affect how stakeholders, including citizens, practitioners,

and commissioners, perceive AI integration. Ownership structures will also affect incentives surrounding profit-making and the sharing of best practice, thereby affecting dynamics of commercialisation in CAs.

A related issue concerns ethics. It will be particularly important to assess the privacy implications of AI integration into democratic innovations, and evaluate how ethical data governance principles, such as informed consent and data anonymisation, can be applied within democratic innovations to uphold the privacy of citizens. Since ownership structures are likely to affect the data governance approach, such issues could be explored in tandem. This future research would help chart a path towards a political economy for AI integration into democratic innovations that serves the public interest.

### Appendix A – Overview of Key Informants

Name	Experience
Kelly McBride	Deliberative democracy lead at TPXImpact. She was the co-lead designer and facilitator for Scotland's Climate Assembly.
Yago Bermejo	Founder of Deliberativa. He is a world leader in transnational deliberation and was involved in both the Future of Europe Conference and the Global Assembly on the Climate and Ecological Crisis.
Claire Mellier	Knowledge and Practice lead at Iswe Foundation and an experienced facilitator of deliberation. She was the co-initiator of the Global Assembly on the Climate and Ecological Crisis.
Johnny Stormonth-Darling	Programme Manager of Citizens' Assemblies at the Iswe Foundation. He was involved as the note-taker coordinator and hosting coordinator in the Global Assembly on the Climate and Ecological Crisis and co-led the authoring of its report.

### Notes

- <sup>1</sup> CAs tend to be stand alone, non-institutionalised processes. However, some CAs take on permanent models, such as the Ostbelgien model. These permanent models iteratively go through this 'lifecycle'.
- <sup>2</sup> A third necessary dimension for assessing the desirability of AI integration into democratic innovations concerns ethics: namely, can AI be integrated in an ethical manner? There are many important ethical dimensions for consideration, such as privacy, bias, and accountability. Here, I focus on the democratic and institutional implications of AI integration, leaving further ethical considerations to future research.
- <sup>3</sup> Whilst LLMs could play a key role in all of the applications I discuss, two of the more complex applications – AI translation and AI facilitation – may benefit from deploying other AI tools as well, such as neural machine translation and speech-time tracking respectfully.

## Acknowledgements

I would especially like to thank Professor Oliver Escobar for his invaluable support and feedback on this project. I would also like to thank my four key informant interviewees – Kelly McBride, Yago Bermejo, Johnny Stormonth-Darling and Claire Mellier – as well as Lukas Salecker and Flynn Devine, for sharing their valuable insights on this topic. The comments and feedback of two anonymous reviewers were also very enriching to this paper. Lastly, I am very grateful to the Edinburgh Futures Institute for providing a supportive environment for my learning and development throughout my master's programme, enabling this paper to become a reality.

## References

- Aitamurto, T., & Landemore, H. E.** (2015). Five design principles for crowdsourced policymaking: Assessing the case of crowdsourced off-road traffic law in Finland. *Journal of Social Media for Organizations*, 2(1), 1–19.
- Alnemr, N.** (2020). Emancipation cannot be programmed: blind spots of algorithmic facilitation in online deliberation. *Contemporary Politics*, 26(5), 531–552. DOI: <https://doi.org/10.1080/13569775.2020.1791306>
- Asenbaum, H.** (2016). Facilitating inclusion: Austrian wisdom councils as democratic innovation between consensus and diversity. *Journal of Deliberative Democracy*, 12(2). DOI: <https://doi.org/10.16997/jdd.259>
- Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M.** (2018). Deliberative democracy. In *The Oxford handbook of deliberative democracy* (pp. 1–34). Oxford, UK: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780198747369.001.0001>
- Bakker, M., Chadwick, M., Sheahan, H., Tessler, M., Campbell-Gillingham, L., Balaguer, J., McAleese, N., Glaese, A., Aslanides, J., Botvinick, M., & Summerfield, C.** (2022). Fine-Tuning language models to find agreement among humans with diverse preferences. *Advances in Neural Information Processing Systems*, 35, 38176–38189.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S.** (2021, March). On the dangers of stochastic parrots: Can language models be too big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency of the Association for Computing Machinery*. New York, United States (610–623). DOI: <https://doi.org/10.1145/3442188.3445922>
- Benjamin, R.** (2019). *Race after technology: Abolitionist tools for the new Jim Code*. John Wiley & Sons.
- Carson, L.** (2017). *Enhancing citizen jurors' critical thinking capacity* (research and development note). newDemocracy Foundation. Retrieved from <https://www.newdemocracy.com.au/2017/03/17/enhancing-citizen-jurors-critical-thinking-capacity/>
- Chambers, S.** (2022). Methods of theorizing. In S. A. Ercan, H. Asenbaum, N. Curato, & R. F. Mendonça (Eds.), *Research methods in deliberative democracy* (pp. 27–39). Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780192848925.001.0001>
- Chang, S., Ciesla, E., Fishkin, S. J., Kartsang, T., & Siu, A.** (2023). *Metaverse Community Forum: Results Analysis*. Stanford Deliberative Democracy Lab. Retrieved from <https://cddrl.fsi.stanford.edu/publication/metaverse-community-forum-results-analysis>
- Coeckelbergh, M.** (2022). Democracy, epistemic agency, and AI: Political epistemology in times of artificial intelligence. *AI and Ethics*, 1–10. DOI: <https://doi.org/10.1007/s43681-022-00239-4>
- Computational Democracy Project.** (2023). *Integrating polis with citizens assemblies*. Retrieved from <https://compdemocracy.org/polis/book/lottery-selected-assemblies/>
- Craglia, M., Hradec, J., & Troussard, X.** (2020). The big data and artificial intelligence: Opportunities and challenges to modernise the policy cycle. In V. Sucha, & M Sienkiewicz (Eds.), *Science for policy handbook* (96–103). Elsevier. DOI: <https://doi.org/10.1016/B978-0-12-822596-7.00009-7>
- Curato, N., Chalaye, P., Conway-Lamb, W., De Pryck, K., Elstub, S., Morán, A., Oppold, D., Romero, J., Ross, M., Sanchez, E., Sari, N., Stasiak, D., Tilikete, S., Veloso, L., von Schneidemesser, D., & Werner, H.** (2023). *Global assembly on the climate and ecological crisis* (2021 Global Assembly Evaluation Report).
- Curato, N., Farrell, D. M., Geissel, B., Grönlund, K., Mockler, P., Pilet, J. B., Renwick, A., Rose, J., Setälä, M., & Suiter, J.** (2021). Deliberative mini-publics: Core design features. In *Deliberative mini-publics*. Bristol University Press. DOI: <https://doi.org/10.1332/policypress/9781529214109.001.0001>
- Danaher, J.** (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, 29(3), 245–268. DOI: <https://doi.org/10.1007/s13347-015-0211-1>
- de Vassimon Manela, D., Errington, D., Fisher, T., van Breugel, B., & Minervini, P.** (2021, April). Stereotype and skew: Quantifying gender bias in pre-trained and fine-tuned language models. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, online (pp. 2232–2242). DOI: <https://doi.org/10.18653/v1/2021.eacl-main.190>
- Devine, F.** (2023). *How AI could scale citizens' assemblies*. Retrieved from [https://flynndevine.substack.com/p/how-ai-could-make-citizens-assemblies-63a?r=1mzenl&utm\\_campaign=post&utm\\_medium=web](https://flynndevine.substack.com/p/how-ai-could-make-citizens-assemblies-63a?r=1mzenl&utm_campaign=post&utm_medium=web)
- Doerr, N.** (2018). *Political translation: How social movement democracies survive*. Cambridge University Press. DOI: <https://doi.org/10.1017/9781108355087>
- Dryzek, J. S.** (2012). *Foundations and frontiers of deliberative governance*. Oxford University Press.
- Elstub, S., & Escobar, O.** (2019). Defining and typologising democratic innovations. In *Handbook of democratic innovation and governance* (11–31). Cheltenham, UK: Edward Elgar Publishing. DOI: <https://doi.org/10.4337/9781786433862.00009>

- Escobar, O., & Elstub, S.** (2017). *Forms of mini-publics* (research and development note). New Democracy.
- Escobar, O.** (2019). Facilitators: The micropolitics of public participation and deliberation. In *Handbook of democratic innovation and governance* (178–195). Cheltenham, UK: Edward Elgar Publishing. DOI: <https://doi.org/10.4337/9781786433862.00022>
- Fayyad, U. M.** (2023). From stochastic parrots to intelligent assistants—The secrets of data and human interventions. *IEEE Intelligent Systems*, 38(3), 63–67. DOI: <https://doi.org/10.1109/MIS.2023.3268723>
- Feng, Y., Qiang, J., Li, Y., Yuan, Y., & Zhu, Y.** (2023). Sentence simplification via large language models. *arXiv Preprint*. arXiv:2302.11957.
- Fishkin, J., Garg, N., Gelauff, L., Goel, A., Munagala, K., Sakshuwong, S., Siu, A., & Yandamuri, S.** (2018). Deliberative democracy with the online deliberation platform. *Paper presented at the 7th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2019)*. Washington State, United States.
- Çaşıorowska, A.** (2023). Sortition and its Principles: Evaluation of the Selection Processes of Citizens' Assemblies. *Journal of Deliberative Democracy*, 19(1). DOI: <https://doi.org/10.16997/jdd.1310>
- Global Assembly Team.** (2022). *Report of the 2021 Global Assembly on the Climate and Ecological Crisis, Earth*. Retrieved from <http://globalassembly.org>
- Guerreiro, N. M., Alves, D., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F.** (2023). Hallucinations in large multilingual translation models. Preprint. *arXiv*. arXiv:2303.16104. DOI: [https://doi.org/10.1162/tacl\\_a\\_00615](https://doi.org/10.1162/tacl_a_00615)
- Hartmann, J., Schwenzow, J., & Witte, M.** (2023). *The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation*. Preprint. *arXiv*. arXiv:2301.01768. DOI: <https://doi.org/10.2139/ssrn.4316084>
- Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H.** (2023). How good are gpt models at machine translation? A comprehensive evaluation. Preprint. *arXiv*. arXiv:2302.09210.
- Hewer, R., Smith, K., & Fergie, G.** (2019). The social functionality of humor in group-based research. *Qualitative Health Research*, 29(3), 431–444. DOI: <https://doi.org/10.1177/1049732318800675>
- Huang, H., Tang, T., Zhang, D., Zhao, W. X., Song, T., Xia, Y., & Wei, F.** (2023). Not all languages are created equal in llms: improving multilingual capability by cross-lingual-thought prompting. Preprint. *arXiv*. arXiv:2305.07004. DOI: <https://doi.org/10.18653/v1/2023.findings-emnlp.826>
- Jaramillo, M. C., & Steiner, J.** (2014). Deliberative transformative moments: A new concept as amendment to the discourse quality index. *Journal of Deliberative Democracy*, 10(2). DOI: <https://doi.org/10.16997/jdd.210>
- Kordzadeh, N., & Ghasemaghaei, M.** (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. DOI: <https://doi.org/10.1080/0960085X.2021.1927212>
- Landemore, H.** (2022). Can AI bring deliberation to the masses. Retrieved from <https://www.law.nyu.edu/sites/default/files/Helen%20Landemore%20Can%20AI%20bring%20deliberative%20democracy%20to%20the%20masses.pdf>
- Leavy, S.** (2018). *Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning*. Proceedings of the 1st international workshop on gender equality in software engineering, Gothenburg, Sweden (pp. 14–16). DOI: <https://doi.org/10.1145/3195570.3195580>
- Lyu, C., Xu, J., & Wang, L.** (2023). *New trends in machine translation using large language models: Case examples with chatgpt*. Preprint. *arXiv*. arXiv:2305.01181.
- Mialhe, N., & Hodes, C.** (2017). *Making the AI revolution work for everyone* (Report-to-OECD). The Future Society at the Harvard Kennedy School of Government. Cambridge. Retrieved from <https://www.tuftsgloballeadership.org/sites/default/files/images/resourcehttp://ai-initiative.%20org/wp-content/uploads/2017/08/Making-the-AI-Revolution-work-for-everyone.-.-MARCH-2017.%20Pdf.s/Mialhe%20Reading.pdf>
- Nicholas, G., & Bhatia, A.** (2023). *Lost in translation: Large language models in non-English content analysis*. Preprint. *arXiv*. arXiv:2306.07377.
- Roberts, J., & Escobar, O.** (2015). *Involving communities in deliberation: A study of three citizens' juries on onshore wind farms in Scotland*. Retrieved from [https://www.climateexchange.org.uk/media/1438/citizens\\_juries\\_-\\_full\\_report.pdf](https://www.climateexchange.org.uk/media/1438/citizens_juries_-_full_report.pdf)
- Ryan, M., & Smith, G.** (2014). Defining mini-publics. In K. Grönlund, A. Bächtiger, & M. Setälä (Eds.), *Deliberative mini-publics: Involving citizens in the democratic process*. Colchester: Ecpr Press.
- Scottish Learning Disabilities Observatory.** (2021). *Research voices: Including the voices of people with learning disabilities in health research* (Citizens' Jury Evaluation Report). Retrieved from <https://www.sldo.ac.uk/media/1939/rv-evaluation.pdf>
- Setälä, M., & Smith, G.** (2018). Mini-publics and deliberative democracy. In A. Bächtiger, et al (Eds.), *The Oxford handbook of deliberative democracy* (300–314), Oxford Handbooks.
- Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., Suzman, T., Ganguli, D., & Megill, C.** (2023). *Opportunities and risks of LLMs for scalable deliberation with Polis*. Preprint. *arXiv*. arXiv:2306.11932.
- Smith, G.** (2009). *Democratic innovations: Designing institutions for citizen participation*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511609848>
- Sprain, L., Carcasson, M., & Merolla, A. J.** (2014). Utilizing “on tap” experts in deliberative forums: Implications for design. *Journal of Applied Communication Research*, 42(2), 150–167. DOI: <https://doi.org/10.1080/00909882.2013.859292>

- Taylor, G. A. J., & Blake, B. J.** (2015). *Key informant interviews and focus groups* (pp. 153–165). Springer.
- Tucker, C.** (2018). Privacy, algorithms, and artificial intelligence. In *The economics of artificial intelligence: An agenda* (pp. 423–437). University of Chicago Press. DOI: <https://doi.org/10.7208/chicago/9780226613475.003.0017>
- von Eschenbach, W. J.** (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607–1622. DOI: <https://doi.org/10.1007/s13347-021-00477-0>
- Wang, P.** (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. DOI: <https://doi.org/10.2478/jagi-2019-0002>
- Wyss, D., & Beste, S.** (2017). Artificial facilitation: Promoting collective reasoning within asynchronous discussions. *Journal of Information Technology & Politics*, 14(3), 214–231. DOI: <https://doi.org/10.1080/19331681.2017.1338175>
- Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B.** (2023). *Benchmarking large language models for news summarization*. Preprint. *arXiv*. arXiv:2301.13848. DOI: [https://doi.org/10.1162/tacl\\_a\\_00632](https://doi.org/10.1162/tacl_a_00632)

**How to cite this article:** McKinney, S. (2024). Integrating Artificial Intelligence Into Citizens' Assemblies: Benefits, Concerns and Future Pathways. *Journal of Deliberative Democracy*, 20(1), pp. 1–12. DOI: <https://doi.org/10.16997/jdd.1556>

**Submitted:** 18 December 2023      **Accepted:** 16 April 2024      **Published:** 01 July 2024

**Copyright:** © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



*Journal of Deliberative Democracy* is a peer-reviewed open access journal published by University of Westminster Press.