Government Information Quarterly 42 (2025) 102079

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Artificial Intelligence in deliberation: The AI penalty and the emergence of a new deliberative divide

Andreas Jungherr [a,*], Adrian Rauchfleisch [b]

[a] Institute of Political Science, University of Bamberg, Bamberg, Germany
[b] Graduate Institute of Journalism, National Taiwan University, Taipei City, Taiwan

ARTICLE INFO

Keywords:
Deliberation
Artificial intelligence
Survey experiment
Political behavior
Participation

ABSTRACT

Advances in Artificial Intelligence (AI) promise help for democratic deliberation, such as processing information, moderating discussion, and fact-checking. But public views of AI's role remain underexplored. Given widespread skepticism, integrating AI into deliberative formats may lower trust and willingness to participate. We report a preregistered within-subjects survey experiment with a representative German sample ($n = 1850$) testing how information about AI-facilitated deliberation affects willingness to participate and expected quality. Respondents were randomly assigned to descriptions of identical deliberative tasks facilitated by either AI or humans, enabling causal identification of information effects. Results show a clear AI penalty: participants were less willing to engage in AI-facilitated deliberation and anticipated lower deliberative quality than for human-facilitated formats. The penalty shrank among respondents who perceived greater societal benefits of AI or tended to anthropomorphize it, but grew with higher assessments of AI risk. These findings indicate that AI-facilitated deliberation currently faces substantial public skepticism and may create a new "deliberative divide." Unlike traditional participation gaps linked to education or demographics, this divide reflects attitudes toward AI. Efforts to realize AI's affordances should directly address these perceptions to offset the penalty and avoid discouraging participation or exacerbating participatory inequalities.

1. Introduction

Advances in Artificial Intelligence (AI) have sparked interest in its potential to support digital deliberation (Landemore, 2024; Tsai et al., 2024). While digital formats have already expanded the deliberative toolkit (Landemore, 2021), persistent challenges remain, such as surfacing relevant information in large groups, moderating contributions, verifying claims, and summarizing debates. AI offers promising solutions to these challenges. However, for these promises to be realized, we must understand whether and how the integration of AI into deliberative processes shapes public expectations of deliberative quality and whether it affects people's willingness to participate in such formats at all.

AI is a contested technology, evoking both high expectations and deep anxieties. Public attitudes are increasingly polarized regarding its societal risks and benefits (O'Shaughnessy et al., 2023; Zhang, 2024; Zhang & Dafoe, 2019). These ambivalent views raise the possibility that people's views on AI may shape both their willingness to participate in AI-facilitated deliberation and their expectations of deliberative quality for respective formats. Given widespread skepticism toward the use of AI in politics (Jungherr et al., 2024; Rauchfleisch et al., 2025), this could negatively influence attitudes toward AI-facilitated deliberation.

Recent studies suggest that participants sometimes prefer outputs from AI-facilitated deliberation to outputs from human-facilitated alternatives (Tessler et al., 2024). Some interpret this as evidence that AI can enhance deliberative processes. But such conclusions overlook a critical aspect: much of this research occurs under experimental conditions where participation is required or incentivized. In practice, deliberative processes are voluntary. If the prospect of interacting with AI deters potential participants, the benefits of AI may never be realized in practice.

Democratic deliberation is not only about producing efficient outcomes. It is fundamentally about enabling inclusive participation and providing individuals with voice and recognition in collective decision-making (Bächtiger & Dryzek, 2024; Fishkin, 2018; Lafont, 2020; Neblo et al., 2018). If certain groups opt out due to discomfort with AI, this undermines deliberative legitimacy. To assess AI's potential role in democratic deliberation, we must therefore look beyond its technical

* Corresponding author at: University of Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany.
E-mail address: andreas.jungherr@uni-bamberg.de (A. Jungherr).

https://doi.org/10.1016/j.giq.2025.102079
Received 22 March 2025; Received in revised form 16 September 2025; Accepted 16 September 2025
Available online 24 September 2025

performance and consider how people respond to the idea of AI-facilitated interaction. This paper addresses this gap.

We examine whether learning that an online deliberation will be facilitated by AI affects (a) people's willingness to participate and (b) their expectations about deliberative quality. We also ask how these effects are moderated by people's prior attitudes toward AI, especially their assessments of its societal risks and benefits.

To answer these questions, we conducted a preregistered survey experiment with a nationally representative sample of adults in Germany ($n = 1850$). Respondents were presented with descriptions of key tasks in democratic deliberation. Using random assignment, each task was attributed either to a human or an AI system. This design allows us to estimate the causal impact of information about AI facilitation in deliberation on participation intentions and expected deliberative quality.

Our findings reveal a significant AI penalty: respondents were, on average, significantly less willing to participate in deliberations facilitated by AI than in those led by humans. They also rated the expected quality of AI-facilitated deliberation lower. Participation willingness was shaped not only by general predispositions for politics and deliberation, such as political interest and need for cognition, but also by respondents' attitudes toward AI. The AI penalty is smaller among those with positive views of AI's societal benefits or a greater tendency to anthropomorphize AI, and larger among those with greater AI-related risk concerns.

These results highlight a core challenge for AI-facilitated deliberation: public acceptance. Even if AI-enabled systems improve the procedural efficiency or output of deliberation, their success ultimately depends on voluntary participation. Our study also reveals a new form of deliberative divide. Traditional deliberative formats already struggle to engage individuals with low political interest or limited cognitive engagement (Coleman & Blumler, 2009; Landemore, 2020). AI-facilitated formats introduce an additional barrier, one rooted not in political attitudes or social background, but in people's beliefs about AI. Unlike established participation gaps based on education or socio-demographic status, this AI-based divide cuts across these lines. Yet it is no less consequential. As political participation becomes increasingly digital and AI-facilitated, people's comfort with and trust in AI will shape who participates and who does not.

2. Theory

2.1. AI-facilitated deliberation

Advances in generative Artificial Intelligence (AI) have raised awareness of how AI could support democratic deliberation by facilitating tasks in deliberative formats. AI is a broad umbrella term that includes a wide variety of approaches that allow machines "to function appropriately and with foresight in [their] environment" (Nilsson, 2010, p. xiii). Scholars have proposed multiple ways AI might improve the performance and quality of digitally mediated deliberative processes, including both curational/analytic (non-generative) and generative approaches (Landemore, 2024; McKinney, 2024; Small et al., 2023; Tessler et al., 2024; Tsai et al., 2024). For example, we can imagine AI-facilitated deliberation formats that use curational/analytic methods (based on expert or data-driven rules) to surface or hide contributions in large-scale online deliberation, akin to the recommendation algorithms that structure feeds in digital content delivery (Narayanan, 2023). Recent advances in generative AI point to additional uses: by creating or transforming content, generative systems can be integrated into tasks such as summarizing arguments in large-scale deliberation, enabling chatbot-like interrogation of problems and contributions, providing translations in multilingual settings, adjusting contributions for clarity, tone, or persuasiveness, or prompting hesitant participants to engage. Going further, speech or video models can generate audio content or interact with users through life-like avatars. This distinction points to an interesting difference in AI-facilitated deliberation. Curational/analytic (non-generative) approaches primarily structure the information environment (e.g. surfacing, emphasizing, or de-emphasizing contributions) whereas generative assistance creates or transforms content (including participants' contributions and interactions) and, in some implementations, engages via voice or avatar interfaces. These approaches intervene at different points in the process and may differentially affect perceived autonomy and authenticity. The opportunities for AI-facilitated deliberation are many, and we are likely only scratching the surface of potential applications.

Based on this discussion, we can define AI-facilitated deliberation as deliberative formats in which AI is used to fulfill specific tasks supporting deliberative processes and functions. Focusing on specific tasks in deliberation, scholars, developers, and practitioners have discussed the use of AI in deliberation for the summary of information and arguments (Arana-Catania et al., 2021; Bakker et al., 2022; Chowanda et al., 2017; McKinney, 2024; Small et al., 2023; Tessler et al., 2024), to support deliberative exchanges (Arana-Catania et al., 2021; Argyle et al., 2023; Dooling & Febrizio, 2023; Fishkin et al., 2019; McKinney, 2024) by enforcing civility and factuality (Agarwal et al., 2024; Fishkin et al., 2019; Giarelis et al., 2024) and enabling comprehension between diverse sets of participants (Feng et al., 2023; McKinney, 2024; Small et al., 2023), and to support consensus building and decision making by mapping and aggregating opinions and preferences of participants (Arana-Catania et al., 2021; Fish et al., 2024; Fishkin et al., 2019; Gudiño-Rosero et al., 2024; Konya et al., 2023; McKinney, 2024; Small et al., 2023). Using AI for these tasks promises to offset the persisting challenges of participant management and information aggregation in large-scale online deliberation (Landemore, 2024).

However, while the technical potential for AI to improve deliberation seems evident, we do not know how people react to the use of AI in deliberation. Particularly, we do not know if and how the prospect of AI-facilitated deliberation influences people's willingness to participate in deliberation and their expectations of the deliberative quality of AI-facilitated processes.

Technical utility does not automatically lead to public acceptance. Democratic deliberation is not only about finding solutions to collective problems; fundamentally, it is about enabling inclusive participation and providing participants with voice and recognition in the democratic process (Bächtiger & Dryzek, 2024; Fishkin, 2018; Lafont, 2020; Neblo et al., 2018). This is especially relevant in the context of well-known issues with AI, including biases in data-driven learnings, intransparency of alignment decisions interfering with model-driven outputs, and corporate (and often foreign) control of AI models and technology (Barocas et al., 2023; Gabriel, 2020; Jungherr & Rauchfleisch, 2025). These issues in the underlying technology are broadly discussed in the context of AI more general and potentially color people's sense of AI-facilitated deliberation. So, while AI might efficiently support collective decision-making through information processing and facilitating interactions, its role as mediator or moderator of people's contributions during the deliberative process might negatively impact participants' sense of voice and recognition (Alnemr, 2020; Lazar & Manuali, 2024).

Including AI in deliberative processes marks a significant departure from previous waves of digital innovation. Earlier developments, such as the integration of Web 2.0 and social media technologies into digital deliberation platforms, primarily enhanced interactivity and user engagement (Davies et al., 2009; Davies & Chandler, 2012; Rose & Sæbø, 2010; Sæbø et al., 2009; Steibel & Estevez, 2015). These innovations introduced features such as interactive comment threads, multimedia elements like video and voice, and backchannels for informal exchanges. While these additions improved the capabilities of deliberative platforms and deepened opportunities for participant interaction, they did not challenge the underlying structure of human agency in the deliberative process.

By contrast, AI introduces qualitatively different dynamics. Unlike Web 2.0 tools, AI systems may act with a degree of autonomy, serving

not just as tools but as mediators or even interlocutors in deliberative exchanges. This shift implies a redistribution of agency, from human moderators and participants to algorithmic agents. Such a development can generate public skepticism, particularly among participants who value essential human qualities in deliberation and who perceive AI as a technological "other" rather than a legitimate deliberative actor. In other words, willingness to participate in AI-facilitated deliberation is shaped not only by the design of the platform but by participants' underlying views of AI itself.

AI is, therefore, not simply another technical enhancement to deliberative platforms. Its transformative, generative, mediating, and agentic qualities may lead people to experience it as an autonomous, potentially even human-like, presence with influence over the process (Epley et al., 2007; Ikari et al., 2023; Waytz et al., 2010). Explaining participation in AI-facilitated deliberation thus requires more than applying traditional models, such as the Technology Acceptance Model (Davis, 1989) or its numerous derivatives (Choi & Song, 2020; King & He, 2006; Marangunić & Granić, 2015; Oni et al., 2017). It demands attention to participants' specific perceptions of AI and the social meanings they attach to it.

2.2. Explaining public reactions to information about AI-facilitated deliberation

Introducing AI to deliberative formats risks courting controversy. AI is a highly controversial technology with bifurcated expectations (O'Shaughnessy et al., 2023; Zhang, 2024; Zhang & Dafoe, 2019) and strong public skepticism, especially regarding its applications in politics (Jungherr et al., 2024; Rauchfleisch et al., 2025). To understand people's views on AI-facilitated deliberation, we provide an explanatory model that considers their attitudes toward the role of technology in society, specific views on AI and usage experience, as well as traditional explanatory factors of participation in deliberative formats.

AI is an advanced technology with which many people have no direct contact. Therefore, for many people, their assessment of AI will be downstream from their overall assessment of technology and its role in society. Going further, if people express specific assessments of AI risks and benefits for different societal areas, we can expect those to matter for their reaction to AI-facilitated deliberation (Bao et al., 2022; Zhang & Dafoe, 2019). Furthermore, people's direct experience with AI, either at work or in their private life, is also an indicator of their openness toward AI and, therefore, is also an important moderating factor. Finally, we consider the tendency of people to anthropomorphize machines – that is, to attribute human characteristics to them (Epley et al., 2007; Ikari et al., 2023; Waytz et al., 2010). This tendency has been observed in people's increased openness to technology use in other contexts and their decreased criticality toward it. This leads us to expect AI anthropomorphism (Folk et al., 2025; Ibrahim et al., 2025) to figure as a moderator for attitudes toward AI-facilitated deliberation.

More generally, willingness to participate in deliberation has been explained by a combination of material and cognitive factors (Neblo et al., 2010). Drawing on the influential civic voluntarism model (Schlozman et al., 2018; Verba et al., 1995), Neblo et al. (2010) argue that political participation depends on individuals' available resources, particularly time and money, as well as their civic skills and psychological engagement with politics.

Building on this foundation, Neblo et al. (2010) further contend that deliberative participation involves additional considerations beyond those relevant for conventional political acts. Specifically, they highlight the importance of psychological fit with the deliberative process itself, including a willingness to engage in reasoned argument, and individuals' aptitude for specific deliberative formats. These additional factors help explain who is likely to take part in deliberation and why. They should hold irrespective of whether a deliberative format features AI or not.

Building on these considerations, we develop the following theoretical expectations[1]:

1. AI penalty: AI-facilitated deliberation will reduce participants' willingness to participate and lower their expectation of deliberative quality, compared to human-facilitated formats.
2. Conditional moderation: This AI penalty will vary depending on individual experiences and attitudes toward AI and technology more broadly.
3. Persistence of traditional factors: Established predictors of deliberative participation will operate similarly regardless of AI involvement.

3. Data & methods

To test the causal effects of learning about different forms of AI-facilitated deliberation, we conducted a preregistered within-subjects survey experiment with members of an online panel provided by the market and opinion research company Ipsos. We used quotas for age, gender, region, and education to create a sample representative of the German population aged 18 and older (see Supplementary Materials for details on sampling). Our final sample consisted of 1850 participants, after removing 97 participants who failed at least two out of three attention checks.[2] A prior power analysis indicated a power of 1 for 1800 respondents or more for the interaction effects in our experiment.[3] Before running the survey, we registered our research design, analysis plan, and hypotheses about outcomes.[4] The data collection ran between February 11 and February 18, 2025.

We presented respondents with descriptions of a random selection of a set of 12 deliberation tasks. This set of tasks captures different elements that are important to the successful design and execution of deliberation formats. This includes recruitment of members and topic selection (i.e., balanced participation, topic identification), information and argument processing (i.e., content summarization, argument highlighting, argument structuring, opinion aggregation), discussion moderation (i.e., fact checking, disruption detection, tone & authenticity), and facilitation (i.e., gamification, translation support, role-play empathy). Consistent with the distinction introduced above, these tasks can be implemented using curational/analytic (non-generative) methods (e.g., ranking, classification, clustering, preference aggregation) or generative assistance (e.g., summaries, translations, style/clarity edits), and proposed AI-enabled facilitation in deliberative formats often combine both. Our vignettes intentionally abstracted from implementation details and focused on the function of each task rather than the specific underlying technology, so the design does not isolate technology-class effects. See Table 1 for a full list of tasks, their deliberative function, and the respective description wording.

We presented each respondent with six randomly selected deliberation tasks out of a total of 12 possible tasks. We randomly attributed each task to a human or AI facilitator. Each respondent saw three tasks attributed to humans and three tasks attributed to AI. We then measured the effects on two dependent variables: respondents' willingness to participate in deliberation and their expectations of deliberative quality.

[1] These expectations represent a condensed version of our preregistered research questions and hypotheses. Prior to conducting the study, we preregistered a comprehensive set of hypotheses addressing each variable in our explanatory model, see preregistration https://osf.io/aq42e/?view_only=9f19e324effb42d9a5891ceedaf728fa. For clarity and readability, we summarize them here in the form of the three key expectations listed above. The full set of preregistered hypotheses is available in the Supplementary Materials.

[2] See Supplementary Materials for details on exclusions.

[3] For R-code for the power simulation see preregistration.

[4] For preregistration see https://osf.io/aq42e/?view_only=9f19e324effb42d9a5891ceedaf728fa. We only deviate from the preregistration in the measurement of education. Other than preregistered, we count BA attainment or higher as high education.

Table 1
Deliberative tasks.

Type	Task	Description
Recruitment & Selection	Balanced Participation	[AI tools/Human moderators] select participants for online discussions, ensuring demographic diversity and balancing power dynamics to make sure many different views are heard and given equal weight.
	Topic Identification	[AI tools/Human moderators] identify topics for online discussions by looking for trending issues, recurring concerns, and emerging themes, ensuring the topics are relevant and representative of public interests.
Information and argument processing	Content Summarization	[AI tools/Human moderators] can summarize contributions and information to help participants gain an overview and assess the available content.
	Argument Highlighting	[AI tools/Human moderators] can highlight key arguments to promote critical thinking and emphasize essential discussion points.
	Argument Structuring	[AI tools/Human moderators] can map arguments and detect areas of agreement to guide discussions toward collaborative decision-making.
	Opinion Aggregation	[AI tools/Human moderators] can aggregate opinions and preferences from participants to reflect collective decisions effectively and transparently.
Discussion moderation	Fact Checking	[AI tools/Human moderators] can fact-check claims and summarize material to provide participants with accurate and accessible information.
	Disruption Detection	[AI tools/Human moderators] can maintain respect and balance in conversations by flagging disruptive behavior and promoting inclusivity.
	Tone & Authenticity	[AI tools/Human moderators] can analyze sentiment to assess discourse authenticity and use discourse mapping to ensure a plurality of viewpoints and inclusive participation.
Discussion facilitation	Gamification	[AI tools/Human moderators] can emphasize selected arguments or use gamification features to encourage critical thinking and help participants engage more deeply with evidence and reasoning.
	Translation Support	[AI-powered translation tools/Human translators] can help participants engage in discussions across language barriers, making deliberation more inclusive.
	Role-play Empathy	[AI-driven role-playing agents can simulate/Human moderators can provide] alternative perspectives, helping participants critically engage with opposing arguments.

Importantly, respondents in our study did not interact with an AI system or take part in an actual AI-facilitated deliberation. Instead, they were presented with information about deliberative tasks being performed either by AI or by humans. In this respect, our research design differs from that of other recent studies examining people's reactions to direct interaction with AI in deliberative contexts (Tessler et al., 2024). This could be seen as a limitation, since respondents were not able to experience or evaluate AI's actual performance. However, we view this feature as a strength of our approach. It allows us to isolate the effect of information about AI involvement on people's initial willingness to participate in deliberation. Studies that require secured participation in AI-facilitated settings may overlook a critical dimension of the "AI penalty": the possibility that individuals may opt out of deliberation entirely upon learning that AI is involved. Our design is explicitly intended to capture this prior filtering mechanism.

To identify the effects of AI-facilitated deliberation on deliberative intent, we use a variant of an item previously used by Neblo et al. (2010).[5] After showing respondents a description of an AI- or human-enabled deliberation task within a deliberative format, we ask: "If you had the chance to participate in such a session, how interested do you think you would be in doing so?"[6]

To measure the impact of information about AI-facilitated deliberation on people's expected deliberative quality, we construct an index based on responses to five items that capture distinct dimensions of deliberative quality. After presenting respondents with a description of either an AI- or human-enabled task within a deliberative format, we ask them to indicate their agreement or disagreement with the following statements: (1) "This makes it difficult for people like me to get their voices heard [inverted]"; (2) "This makes it easy to identify the best policy solution"; (3) "This ensures a fair discussion and policy process"; (4) "I trust this process"; and (5) "This will give space for a plurality of diverse views on the issue". These items reflect core features of deliberative quality identified in the normative and empirical literature, including inclusive and pluralistic representation of voice, as well as procedural aspects such as fairness, the capacity to identify appropriate solutions, and participants' trust in the process (Dryzek, 2002; Habermas, 1992; Kingwell, 1995; Niemeyer et al., 2024; Steenbergen et al., 2003; Young, 2002).

Our independent variables are a set of established and new measurements. Following Neblo et al. (2010) and Schlozman et al. (2018), we account for different types of resources influencing people's readiness to participate in deliberation. Money and time available for politics should most strongly influence people's willingness to participate. Income is comparatively easy to measure, but time is more difficult to capture. We account for people's full-time employment status and the number of young children in a household. Another set of resources accounts for the necessary civic skills enabling people to participate; this includes educational attainment as a proxy for early exposure to civic skills and its necessary building blocks and prior experience with political participation. To account for the specifics of political deliberation, we measure people's prior experience with political deliberation. To identify psychological involvement with politics, we consider respondents' political interest, political efficacy, and their strength of party identification. Of the potential measures accounting for psychological aptitude suggested by Neblo et al. (2010), we consider the need for cognition.

We measure the need for cognition with items proposed and validated by Matthes (2006). We asked respondents whether they agreed or disagreed with the following statements: (1) "I find satisfaction in deliberating hard and for long hours"; (2) "The notion of thinking abstractly is appealing to me;" (3) "I really enjoy a task that involves coming up with new solutions to problems;" and (4) "I like tasks that require much thought and mental effort."

We are also interested in how people's general sense of the benefits and risks of technology influences their perceptions of AI-facilitated deliberation. We designed two new scales to measure people's sense of technology's impact on their daily lives.[7] To assess people's sense of technological benefits, we asked for their agreement or disagreement

[5] For a full list of all measures including variables, question wordings, answer options, and key diagnostics see Supplementary Materials on Measures.

[6] Here, we present English translations of the questions we ran in our questionnaire in German. For the original German wording of our questionnaire see our preregistration: https://osf.io/aq42e/?view_only=9f19e324effb42d9a5891ceedaf728fa.

[7] For a validation of both scales see Supplementary Materials on Validation.

with the following three statements: (1) "I feel more in control of my daily activities thanks to technology;" (2) "Technology helps me manage my time more effectively;" and (3) I feel more productive in both my personal and professional life because of technology. To measure people's sense of technology risks, we asked for their agreement or disagreements with the following three statements: (1) "Technology increases my stress levels rather than making life easier;" (2) "Technology reduces the quality of face-to-face interactions with others;" and (3) "Technology contributes to distractions and reduces my ability to focus."

Moving closer to our dependent variables, we ask people for their general sense of AI benefits and risks. We measure this by two lists of items that address perceived benefits and risks of AI in different societal areas, including the economy, international security, state capacity, election campaigns, and existential threats. We selected these areas by drawing on research on AI risk and benefit perceptions (Bao et al., 2022; Zhang & Dafoe, 2019) and attitudes toward technology (Binder et al., 2012; Siegrist & Visschers, 2013). By querying respondents for their agreement and disagreement with these statements, we develop an index of their general attitudes toward the benefits and risks of AI.

To account for respondents' experience with AI, we ask how frequently they are using AI-supported applications in their (1) "professional or work environment" and (2) "personal and spare time". We combine the two items into a mean index.

Finally, to measure respondents' inclination to project human faculties onto AI, we build on items proposed by Folk et al. (2025). From their original measurement scale, we identified four items with high reliability, adjusted their wording slightly, and used them to build our index of AI anthropomorphism. Our adjusted items are: (1) "AI has the potential to develop a sense of humor;" (2) "An AI can have a unique personality;" (3) "AI has the capacity to develop consciousness;" and (4) "AI can develop a sense of morals and ethics."

In Table 2, we document descriptive statistics for dependent and independent variables, linked to preregistered research questions and hypotheses.

We estimated two multilevel models with varying intercepts for participants and cases to test our research questions and hypotheses (see Supplementary Materials for additional model checks). We use the varying intercepts to account for the nested data. Each participant saw six different tasks out of a total of twelve tasks. We have between 461 and 464 observations for each specific task and condition (X^2 (11, N = 11,000) = 0.01, p = 1). Thus, the total number of observations for the models is 11,000. We estimated two models, one for each dependent variable, that included all the independent variables and the preregistered covariates age and gender (for the complete models and details on data imputation, see Supplementary Materials). The independent variables used as interaction terms (conditional moderation) were all mean-centered before including them in the model. Furthermore, as specified in the preregistration, we performed data imputation for predictors that allowed non-responses (i.e., income or employment status) by following the procedures recommended in the literature (van Buuren, 2018; van Buuren & Groothuis-Oudshoorn, 2011).

To better understand the motivations behind low interest in AI-facilitated participation, we conducted an additional exploratory analysis using responses to an open-ended question presented at the end to respondents who had indicated at least once that they had no interest in participating in such a format (by selecting 1 or 2 on a seven-point scale). The responses were automatically categorized using OpenAI's 'gpt-4o-mini-2024-07-18' model (details in the Supplementary Materials).

Table 2
Descriptive statistics for dependent and independent variables.

Variable	M (SD)	n
Willingness to participate in deliberation	3.95 (1.98)	11,100
Deliberative quality (5 items, α = 0.76–0.84)	4.16 (1.29)	11,100
Technology benefits (3 items, α = 0.86)	4.63 (1.38)	1850
Technology risks (3 items, α = 0.76)	3.65 (1.43)	1850
AI benefits (5 items, α = 0.87)	3.76 (1.35)	1850
AI risks (5 items, α = 0.83)	4.58 (1.33)	1850
AI experience (2 items, α = 0.77)	2.72 (1.66)	1850
AI anthropomorphism (4 items, α = 0.88)	3.18 (1.54)	1850
Income	8.52 (2.85)	1716
Employment (full-time = 1)	0.49	1823
Children, 12 or younger	0.27 (0.82)	1721
Education (University degree incl. Bachelor = 1)	0.23	1850
Past participatory activity	0.68 (1.08)	1641
Past deliberative activity	0.39 (0.90)	1715
Political interest	4.91 (1.77)	1850
Political efficacy (2 items, Spearman-Brown = 0.67)	3.34 (1.57)	1850
PID strength	4.55 (1.95)	1850
Need for cognition (4 items, α = 0.9)	4.72 (1.38)	1850

4. Results

4.1. Public skepticism toward AI in deliberation: evidence of an AI penalty in participation and quality assessments

An important challenge for proponents of AI-facilitated deliberative processes is public skepticism: Do citizens respond differently when deliberative tasks are facilitated by AI instead of humans? And if so, does this skepticism reduce their willingness to participate or undermine perceived deliberative quality?

To test this, we conducted an exploratory analysis comparing public reactions to tasks facilitating digital deliberation when described as being performed by either AI or human actors. As shown in Fig. 1, we plot respondents' willingness to participate and their expectation of deliberative quality across various tasks, each randomly attributed to an AI or a human facilitator.[8]

The pattern shown in Fig. 1 is clear: respondents consistently express a lower willingness to engage in deliberation and assign lower quality ratings when key tasks (such as content summarization, argument structuring, or opinion aggregation) are performed by AI rather than humans. Notably, these tasks are precisely those that scholars have identified as promising areas for AI support in deliberation (Berliner, 2024; Landemore, 2024; Tessler et al., 2024; Tsai et al., 2024). This divergence suggests that while experts emphasize AI's potential for enhancing information processing, the public remains wary of AI involvement in these functions.

We can estimate the average treatment effect of AI versus human facilitation across all tasks. AI-facilitated deliberation is associated with significantly reduced willingness to participate (b = −0.30, 95 % CI [−0.34, −0.26], $p < 0.001$). When dichotomizing the outcome variable at the midpoint of the scale, we observe a 7.4 percentage point drop in participants indicating a willingness to engage, from 47.44 % (human) to 40.02 % (AI). Similarly, perceived deliberative quality is significantly lower for AI-facilitated formats (b = −0.29, 95 % CI [−0.32, −0.26], p < 0.001), with the share of participants rating quality above the midpoint dropping nearly 10 percentage points (from 56.90 % to 47.53 %).

These patterns point to a robust AI penalty: Across both participation and perceived quality, participants systematically rate AI-facilitated formats more negatively than human-led ones.

To understand the reasons behind this penalty, we analyzed open-ended responses from those who showed no interest in participating in AI-facilitated deliberation for at least one task. Fig. 2 summarizes these coded reasons.

The results reinforce the AI penalty interpretation: 45 % of respondents cited distrust or fear of AI, while 13 % explicitly preferred human interaction. These concerns are not peripheral; they are central

[8] Values are reported after alpha correction (Benjamini & Hochberg, 1995). Stars indicate significant differences identified through *t*-tests.

Fig. 1. Reaction to deliberative tasks performed by AI or humans. Points show means with 95 % confidence intervals; stars indicate significant mean difference after correcting alpha values with the Benjamini-Hochberg procedure.

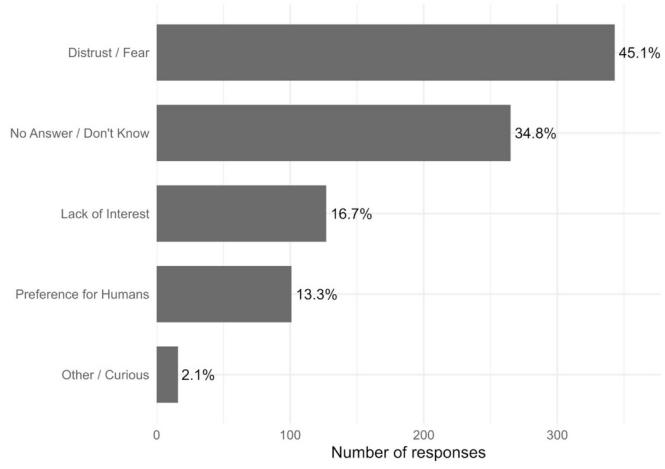

Fig. 2. Reasons for not being willing to participate in AI-facilitated deliberative formats. The percentage is based on the total number of respondents (761) who indicated at least once no interest in participating (selecting 1 or 2 on a seven-point scale).

to respondents' refusal to engage with AI-facilitated deliberation.

Together, these findings strongly support the AI penalty hypothesis: AI-facilitated formats significantly reduce both willingness to participate and expected deliberative quality, primarily due to fundamental concerns about AI trustworthiness and desirability as a partner in deliberation.

4.2. Individual attitudes toward AI moderate the AI penalty in deliberation

A key challenge of deliberative formats lies in the recruitment of participants who reflect the diversity of the general public. Prior research has shown that individuals with lower levels of political engagement and fewer socio-economic resources are typically less likely to participate in deliberative processes (Jacobs et al., 2009; Neblo et al., 2010). The use of AI in deliberation may further strengthen participatory divides. If people's willingness to participate in or positively evaluate AI-facilitated deliberation depends on their prior attitudes toward technology and AI, such systems may inadvertently exacerbate participation biases and undermine perceived legitimacy. To assess this, we test whether the AI penalty is conditional on individual-level variation in attitudes toward AI and technology more broadly.

We begin by examining whether general attitudes toward technology moderate the effect of AI-facilitated versus human-facilitated deliberation. In a preregistered model,[9] we interact our treatment variable (AI vs. human task performance) with measures of perceived technology benefits and technology risks. As shown in Fig. 3, we find no evidence that general technology attitudes moderate treatment effects. Interaction terms for both technology benefit and technology risk are close to zero and statistically insignificant for both outcome measures: willingness to participate (benefit: b = 0.00, 95 % CI [−0.03, 0.04], $p = 0.915$; risk: b = 0.00, 95 % CI [−0.04, 0.03], $p = 0.861$) and deliberative quality expectations (benefit: b = −0.01, 95 % CI [−0.04, 0.02], $p = 0.430$; risk: b = 0.00, 95 % CI [−0.02, 0.03], $p = 0.843$). These null

[9] Fig. 3 documents coefficients to variables linked to theoretically informed preregistered hypotheses. Full model specifications are available in Supplementary Materials on Model Results.

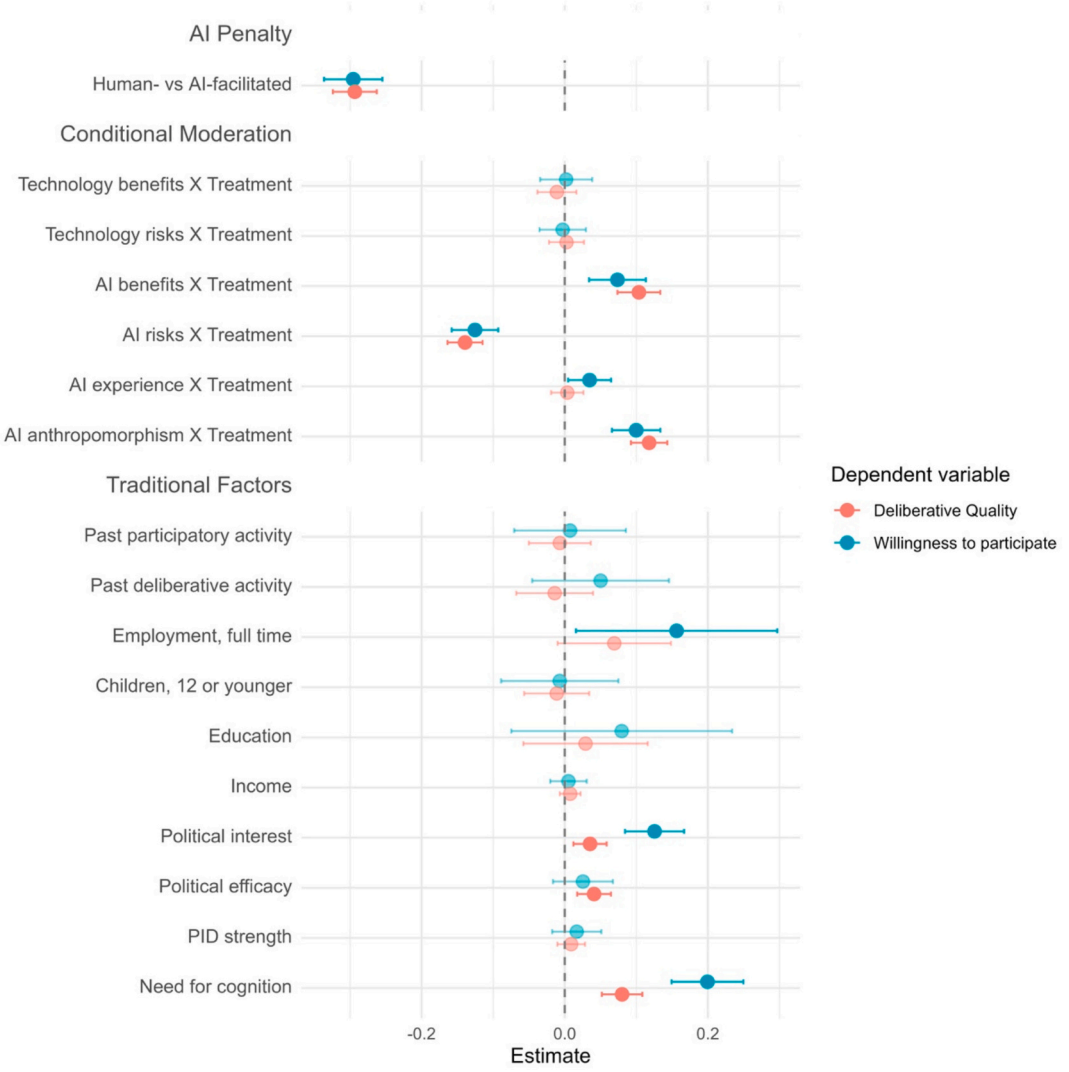

Fig. 3. Models explaining willingness to participate in deliberation and quality expectations.

results suggest that broad attitudes toward technology do not systematically shape how individuals respond to AI-facilitated deliberation.

In contrast, attitudes and experiences specifically related to AI do condition individuals' responses to AI-facilitated deliberative formats. As shown in Fig. 4, respondents with more positive perceptions of AI benefits report significantly higher willingness to participate in AI-facilitated deliberation compared to those informed about human-facilitated formats (b = 0.07, 95 % CI [0.03, 0.11], $p < 0.001$). Conversely, respondents with heightened AI risk perceptions express significantly lower willingness to participate when deliberation is described as AI-facilitated (b = −0.13, 95 % CI [−0.16, −0.09], p < 0.001). Additional interactions further support the conditional nature of the AI penalty: individuals with prior experience using AI (b = 0.03, 95 % CI [0.00, 0.06], $p = 0.023$) and those more inclined to anthropomorphize AI (b = 0.10, 95 % CI [0.07, 0.13], $p < 0.001$) also exhibit greater willingness to participate in AI-facilitated formats.

Fig. 5 displays analogous patterns for expected deliberative quality. High AI benefit assessments are associated with more favorable expectations of AI-facilitated deliberation relative to human-facilitated formats (b = 0.10, 95 % CI [0.07, 0.13], $p < 0.001$), while high AI risk assessments are linked to more critical evaluations (b = −0.14, 95 % CI [−0.16, −0.11], p < 0.001). The tendency to anthropomorphize AI also moderates quality assessments in the expected direction (b = 0.12, 95 % CI [0.09, 0.14], p < 0.001). However, prior AI experience does not significantly moderate deliberative quality expectations (b = 0.00, 95 % CI [−0.02, 0.03], $p = 0.759$).

Taken together, these findings provide robust evidence for conditional moderation of the AI penalty. Individuals who are more skeptical of AI (those perceiving fewer benefits, greater risks, tend not to anthropomorphize AI, or lacking AI-related experience) are more likely to penalize AI-facilitated deliberation, both in terms of participation and expected quality. In contrast, individuals who are more favorable toward, or familiar with AI, or tend to anthropomorphize AI, treat AI-facilitated deliberation similarly to or even more positively than human-facilitated formats. This heterogeneity underscores the socially stratifying potential of AI deployment in democratic processes and highlights the importance of considering attitudinal variation in the design and communication of AI-supported deliberative formats.

4.3. Cognitive engagement, not resources, predicts willingness to participate in deliberation

To evaluate whether established predictors of deliberative participation operate similarly regardless of AI involvement, we draw on the civic voluntarism model (Schlozman et al., 2018; Verba et al., 1995) and prior work on psychological predispositions toward deliberation (Neblo et al., 2010).

We first examine traditional resource-based predictors. As shown in

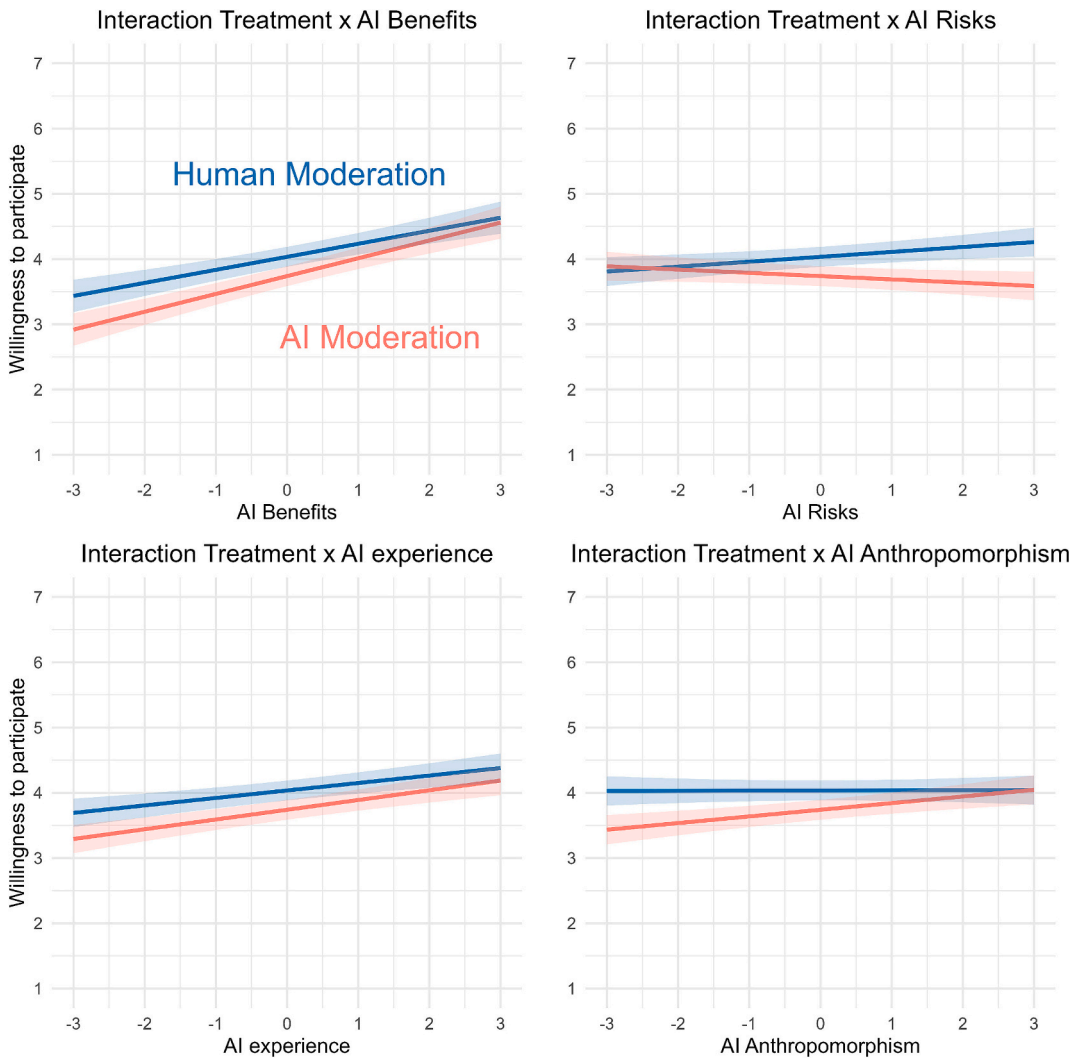

Fig. 4. Interaction effects on interest to participate in deliberation.

Fig. 3, individuals' income (b = 0.01, 95 % CI [−0.02, 0.03], $p = 0.692$), number of young children (b = −0.01, 95 % CI [−0.09, 0.07], $p = 0.868$), educational attainment (b = 0.08, 95 % CI [−0.07, 0.23], $p = 0.312$), and prior participation in political or deliberative activities show no significant effects. Contrary to expectations, full-time employment is associated with greater willingness to participate (b = 0.16, 95 % CI [0.02, 0.30], $p = 0.030$). Together, these results suggest that material constraints (such as time, money, and formal education) do not significantly influence expressed willingness to participate in AI- or human-facilitated deliberation. One explanation may be that these resource considerations are more salient when people must act on their willingness, rather than when reporting it in hypothetical scenarios.

In contrast, cognitive engagement variables perform more consistently. Political interest has a positive and significant association with willingness to participate (b = 0.13, 95 % CI [0.08, 0.17], p < 0.001), while political efficacy (b = 0.03, $p = 0.234$) and party identification strength (b = 0.02, $p = 0.340$) do not. Additionally, in line with expectations, individuals higher in need for cognition are more willing to engage (b = 0.20, 95 % CI [0.15, 0.25], p < 0.001).

These findings indicate that traditional cognitive predictors of deliberative participation remain operative in the context of AI-facilitated formats. Resource-based predictors, by contrast, appear less consequential, potentially due to the hypothetical nature of our participation measure.

5. Discussion

We tested the effects of informing people about the use of AI for performing important tasks within deliberative formats. Our findings showed a clear AI penalty. Respondents expressed less willingness to participate in AI-facilitated deliberation and lower expectations of deliberative quality, compared to human-facilitated deliberation. This AI penalty is especially pronounced for people who are skeptical about AI's role in society. Conversely, for people who tended to see societal benefits in AI, low AI risks, and who anthropomorphized AI, we found little to no AI penalty.

These findings are important as they point to currently little discussed risks to the integration of AI in deliberative formats. If AI reduces people's willingness to participate in deliberative formats, the potentially helpful aspects of AI to the conduct of digitally mediated deliberation are offset. Looking only at potential efficiency gains within deliberative formats while ignoring the attitudinal and motivational effects of AI-facilitated deliberation risks missing an important factor. This is especially true if, as we show, the AI penalty is not uniformly distributed through society but instead follows prior attitudes toward AI. This threatens to introduce a new dimension of deliberative inequality by having people skeptical about AI to withdraw from AI-facilitated deliberative and participatory opportunities. Proponents of AI-facilitated deliberation and AI use in democratic practice in general need to take this risk seriously and propose mechanisms to mitigate the

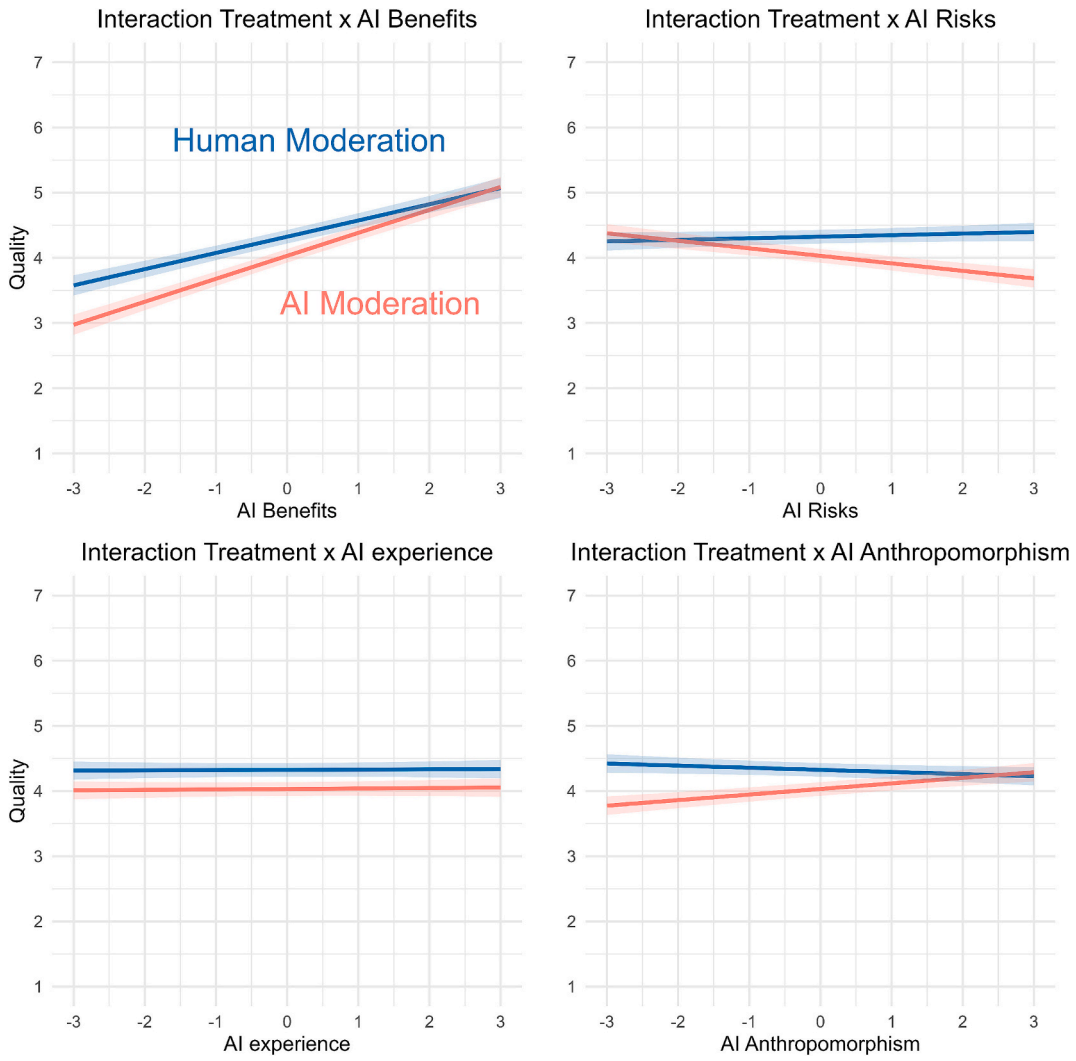

Fig. 5. Interaction effects on expectations of deliberative quality.

new technology-driven participatory divide.

Our findings also inform discussions about AI-facilitated deliberation and participation more broadly. In explaining people's willingness to participate in deliberative formats – AI-facilitated or reliant on humans – factors based on the civic voluntarism model (Schlozman et al., 2018), accounting for material resources did not contribute to the explanation as expected. In contrast, we found that the cognitive drivers of unequal participatory interest held as expected. As expected, both political interest (Schlozman et al., 2018) and need for cognition (Neblo et al., 2010) explained willingness to participate in deliberation. One reason for the surprising underperformance of material factors could be that we only queried people for their interest in participation and did not measure their actual follow-through. The scarcity of material resources, such as time or money, may only come into play when people have to actually spend time on deliberation formats, rather than playing a significant role in an earlier stage of hypothetical commitment (Neblo et al., 2010). Future research could further test this by using behavioral measures. For example, observing whether participants accept an actual invitation to join a deliberative format would allow us to test whether the AI penalty we observe in our study diminishes when participants are given more contextual information about AI facilitation and its efficiency.

We also found, somewhat surprisingly, that in their responses to information about AI-facilitated deliberation for tasks associated with information and argument processing, people appeared especially hesitant. This stands in contrast with the discussion among academics and practitioners, who tend to see the strongest contribution of AI to deliberation in information processing and argument mapping (Landemore, 2024; Tessler et al., 2024; Tsai et al., 2024). While people were skeptical of AI for deliberative information processing and argument mapping, for fact-checking, they rated it similar to human facilitators. This is interesting, as arguably AI, at least in its current state, should be less suited to tasks of fact identification and correction than information processing or argument mapping. This suggests that academics and practitioners should not only focus on technically evaluating AI's information processing and argument mapping tasks, but also consider people's reactions and assessments. Views of academics, practitioners, and the public on the benefits and risks of AI in deliberation might deviate more strongly than academics might think.

Even more fundamentally, the identification of an AI penalty in deliberation contradicts other recent findings that suggest people are open to AI-facilitated deliberation and express higher satisfaction with AI facilitation than with human facilitation (Tessler et al., 2024). One reason for this discrepancy could be that for our respondents, AI-facilitated interventions remain hypothetical, as they do not directly experience these interventions but are only informed about them. Indeed, experiencing AI might make people more supportive of it. But our findings indicate that we should not take AI acceptance as a given. When asked about their interest in participating in deliberation, most people will not yet have direct experience with AI in deliberative formats. The initial skepticism identified by us has, therefore, high

ecological validity as the information our respondents received corresponds with that of regular people when deciding on whether to participate in AI-facilitated deliberation or not. And here, the AI penalty is real. If we are interested in increasing participation in deliberative formats, we must not only focus on people who are already willing to participate. The key challenge lies in including people who are not initially interested in participating. And here our findings indicate that AI-facilitated deliberation might increase hesitancy to participate along new attitudinal dimensions and therefore weaken the representativeness and legitimacy of deliberative formats.

Of course, our study comes with limitations. For one, our findings are based on German respondents. While general psychological processes and information effects can be expected to be similar across countries with comparable cultural, political, and technological constellations, it is unclear whether this is also the case for AI. Countries vary regarding their cultural affinity toward AI (Howell, 2025; Wang, 2024). Accordingly, the penalty identified by us might also vary. Future research should focus on establishing comparative evidence on the AI penalty.

AI-enabled systems facilitating deliberation vary in how they operate. Above, we distinguish between curational/analytic (non-generative) approaches (e.g. ranking, moderating, and mapping contributions via classification, clustering, or preference aggregation) and generative assistance, which creates or transforms content (e.g., summaries, translations, style/clarity edits). These approaches intervene at different points in the process and may differentially affect perceived autonomy and authenticity. Our design does not isolate technology-class effects, so we cannot attribute the observed AI penalty to one mechanism rather than another. Because real platforms often combine both (e.g., analytic ranking with generative summaries), future work should unpack mechanisms and test whether approach (e.g. curational/analytic vs. generative) and specific design choices (e.g. greater transparency, meaningful user control/opt-in, auditability, data minimization) exacerbate or mitigate skepticism. Such evidence would help determine how to design AI-facilitated deliberation to broaden participation and improve expected deliberative quality, rather than inadvertently suppressing willingness to take part.

There are also potential temporal limitations to our study. We show that willingness to participate in and quality expectations of AI-facilitated deliberation vary based on attitudes toward AI. But how these attitudes emerge and how they are distributed across societies, we do not know. Nor do we know the prospective development of public opinion on AI. AI attitudes could turn sour over time, once job replacement or runaway technological change is more clearly felt across societies. Alternatively, AI attitudes could turn more positive once people experience AI-driven increases in welfare and quality of life. Or AI could normalize. This would render AI-facilitated deliberation no longer special, and in turn, lead to AI attitudes playing a lesser role in readiness to participate in AI-facilitated deliberation. Potential temporal variations should, therefore, also be considered by future research.

Overall, we see that attitudes toward AI matter for both willingness to participate and expectations of deliberative quality once AI is introduced in deliberative formats. This is important to keep in mind for designers of deliberative environments so that they do not accidentally introduce new dimensions of unequal participation or quality assessment to a process that already suffers from challenges of unequal recruitment and associated legitimacy challenges based on unequal representation. AI might make digitally mediated deliberation more efficient, but it carries new challenges of public trust and willingness to engage. This challenge is likely to increase if public skepticism toward AI rises.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT 4o for language editing. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Andreas Jungherr: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Adrian Rauchfleisch:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Code availability

Replication code and data are available at the project's OSF repository:

- Replication code and data: Available on OSF https://osf.io/5346r/files/osfstorage?view_only=e204fb9a9d6e4af187c7d95934f92cf0

Funding sources

A. Jungherr's contribution is funded by the European Union under grant agreement No. 101178806. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. A. Rauchfleisch's work was supported by the National Science and Technology Council, Taiwan (R.O.C) (Grant No 113–2628-H-002-018- and Grant No 114-2628-H-002-007-).

Declaration of competing interest

A. Jungherr and A. Rauchfleisch declare no conflict of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

A. Jungherr and A. Rauchfleisch contributed equally to the project. The experiments were preregistered at OSF https://osf.io/aq42e/?view_only=9f19e324effb42d9a5891ceedaf728fa and approved by the IRB at University of Bamberg. Correspondence and requests for materials should be addressed to A. Jungherr.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.giq.2025.102079.

Data availability

Preregistration information and data are available at the project's OSF repository:

- Preregistration: https://osf.io/aq42e/?view_only=9f19e324effb42d9a5891ceedaf728fa
- Data: https://osf.io/5346r/files/osfstorage?view_only=e204fb9a9d6e4af187c7d95934f92cf0

References

Agarwal, D., Shahid, F., & Vashistha, A. (2024). Conversational agents to facilitate deliberation on harmful content in WhatsApp groups. In J. Nichols (Ed.), *vol. 8. Proceedings of the ACM on human-computer interaction* (pp. 1–32). ACM. https://doi.org/10.1145/3687030.

Alnemr, N. (2020). Emancipation cannot be programmed: Blind spots of algorithmic facilitation in online deliberation. *Contemporary Politics, 26*(5), 531–552. https://doi.org/10.1080/13569775.2020.1791306

Arana-Catania, M., Lier, F.-A. V., Procter, R., Tkachenko, N., He, Y., Zubiaga, A., & Liakata, M. (2021). Citizen participation and machine learning for a better democracy. *Digital Government: Research and Practice, 2*(3), 1–22. https://doi.org/10.1145/3452118

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., … Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences, 120*(41), Article e2311627120. https://doi.org/10.1073/pnas.2311627120

Bächtiger, A., & Dryzek, J. S. (2024). *Deliberative democracy for diabolical times: Confronting populism, extremism, denial, and authoritarianism*. Cambridge University Press. https://doi.org/10.1017/9781009261845

Bakker, M. A., Chadwick, M. J., Sheahan, H. R., Tessler, M. H., Campbell-Gillingham, L., Balaguer, J., … Summerfield, C. (2022). Fine-tuning language models to find agreement among humans with diverse preferences. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Vol. 35. NeurIPS 2022: Advances in neural information processing systems 35* (pp. 38176–38189). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/f978c8f3b5f399cae464e85f72e28503-Paper-Conference.pdf.

Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., … A, M. A. X. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior, 130*(107182), 1–10. https://doi.org/10.1016/j.chb.2022.107182

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. The MIT Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berliner, D. (2024). What AI can't do for democracy. *Boston Review*. https://www.bostonreview.net/articles/what-ai-cant-do-for-democracy.

Binder, A. R., Cacciatore, M. A., Scheufele, D. A., Shaw, B. R., & Corley, E. A. (2012). Measuring risk/benefit perceptions of emerging technologies and their potential impact on communication of public opinion toward science. *Public Understanding of Science, 21*(7), 830–847. https://doi.org/10.1177/0963662510390159

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press (Original work published 2012).

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Choi, J.-C., & Song, C. (2020). Factors explaining why some citizens engage in e-participation, while others do not. *Government Information Quarterly, 37*(4), Article 101524. https://doi.org/10.1016/j.giq.2020.101524

Chowanda, A. D., Sanyoto, A. R., Suhartono, D., & Setiadi, C. J. (2017). Automatic debate text summarization in online debate forum. *Procedia Computer Science, 116*, 11–19. https://doi.org/10.1016/j.procs.2017.10.003

Coleman, S., & Blumler, J. G. (2009). *The internet and democratic citizenship: Theory, practice and policy*. Cambridge University Press. https://doi.org/10.1017/CBO9780511818271

Davies, T., & Chandler, R. (2012). Online deliberation design: Choices, criteria, and evidence. In T. Nabatchi, J. Gastil, M. Leighninger, & G. M. Weiksner (Eds.), *Democracy in motion: Evaluating the practice and impact of deliberative civic engagement* (pp. 103–132). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199899265.003.0006.

Davies, T., O'Connor, B., Cochran, A., Effrat, J. J., Parker, A., Newman, B., & Tam, A. (2009). An online environment for democratic deliberation: Motivations, principles, and design. In T. Davies, & S. P. Gangadharan (Eds.), *Online deliberation: Design, research, and practice* (pp. 275–292). University of Chicago Press.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly, 13*(3), 319–420. https://doi.org/10.2307/249008

Dooling, B. C. E., & Febrizio, M. (2023). *Robotic rulemaking*. Brookings Institution. https://www.brookings.edu/articles/robotic-rulemaking/.

Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press. https://doi.org/10.1093/019925043X.001.0001

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Feng, Y., Qiang, J., Li, Y., Yuan, Y., & Zhu, Y. (2023). *Sentence simplification via large language models*. arXiv. https://doi.org/10.48550/arXiv.2302.11957.

Fish, S., Gölz, P., Parkes, D. C., Procaccia, A. D., Rusak, G., Shapira, I., & Wüthrich, M. (2024). Generative social choice. In D. Bergemann, R. Kleinberg, & D. Saban (Eds.), *EC '24: Proceedings of the 25th ACM conference on economics and computation* (p. 985). ACM. https://doi.org/10.1145/3670865.3673547.

Fishkin, J. S. (2018). *Democracy when the people are thinking: Revitalizing our politics through public deliberation*. Oxford University Press. https://doi.org/10.1093/oso/9780198820291.001.0001

Fishkin, J. S., Garg, N., Gelauff, L., Goel, A., Munagala, K., Sakshuwong, S., … Yandamuri, S. (2019). Deliberative democracy with the online deliberation platform. In *HCOMP 2019: The 7th AAAI conference on human computation and crowdsourcing*. HCOMP. https://www.humancomputation.com/2019/assets/papers/144.pdf.

Folk, D., Wu, C., & Heine, S. (2025). Cultural variation in attitudes towards social chatbots. *Journal of Cross-Cultural Psychology, 56*(3), 219–239. https://doi.org/10.1177/00220221251317950

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30*(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Giarelis, N., Mastrokostas, C., & Karacapilidis, N. (2024). A unified LLM-KG framework to assist fact-checking in public deliberation. In A. Hautli-Janisz, G. Lapesa, L. Anastasiou, V. Gold, A. De Liddo, & C. Reed (Eds.), *DELITE 2024: Proceedings of the first workshop on language-driven deliberation technology* (pp. 13–19). ELRA; ICCL. https://aclanthology.org/2024.delite-1.2/.

Gudiño-Rosero, J., Grandi, U., & Hidalgo, C. A. (2024). Large language models (LLMs) as agents for augmented democracy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 382*(2285), 1–17. https://doi.org/10.1098/rsta.2024.0100

Habermas, J. (1992). *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. Suhrkamp.

Howell, B. (2025). WEIRD? Institutions and consumers' perceptions of artificial intelligence in 31 countries. *AI & Society*. https://doi.org/10.1007/s00146-025-02217-w

Ibrahim, L., Akbulut, C., Elasmar, R., Rastogi, C., Kahng, M., Morris, M. R., … Weidinger, L. (2025). *Multi-turn evaluation of anthropomorphic behaviours in large language models*. arXiv. https://doi.org/10.48550/arXiv.2502.07077.

Ikari, S., Sato, K., Burdett, E., Ishiguro, H., Jong, J., & Nakawake, Y. (2023). Religion-related values differently influence moral attitude for robots in the United States and Japan. *Journal of Cross-Cultural Psychology, 54*(6–7), 742–759. https://doi.org/10.1177/00220221231193369

Jacobs, L. R., Cook, F. L., & Delli Carpini, M. X. (2009). *Talking together: Public deliberation and political participation in America*. The University of Chicago Press.

Jungherr, A., & Rauchfleisch, A. (2025). *What do people expect from artificial intelligence? Public opinion on alignment in AI moderation from Germany and the United States*. arXiv. https://doi.org/10.48550/arXiv.2504.12476.

Jungherr, A., Rauchfleisch, A., & Wuttke, A. (2024). *Artificial intelligence in election campaigns: Perceptions, penalties, and implications*. arXiv. https://doi.org/10.48550/arXiv.2408.12613.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740–755. https://doi.org/10.1016/j.im.2006.05.003

Kingwell, M. (1995). *A civil tongue: Justice, dialogue, and the politics of pluralism*. Pennsylvania State University Press.

Konya, A., Schirch, L., Irwin, C., & Ovadya, A. (2023). *Democratic policy development using collective dialogues and AI*. arXiv. https://doi.org/10.48550/arXiv.2311.02242.

Lafont, C. (2020). *Democracy without shortcuts: A participatory conception of deliberative democracy*. Oxford University Press. https://doi.org/10.1093/oso/9780198848189.001.0001

Landemore, H. (2020). *Open democracy: Reinventing popular rule for the twenty-first century*. Princeton University Press.

Landemore, H. (2021). Open democracy and digital technologies. In L. Bernholz, H. Landemore, & R. Reich (Eds.), *Digital technology and democratic theory* (pp. 62–89). The University of Chicago Press. https://doi.org/10.7208/9780226748603-003.

Landemore, H. (2024). Can artificial intelligence bring deliberation to the masses? In R. Chang, & A. Srinivasan (Eds.), *Conversations in philosophy, law, and politics* (pp. 39–69). Oxford University Press. https://doi.org/10.1093/oso/9780198864523.003.0003.

Lazar, S., & Manuali, L. (2024). *Can LLMs advance democratic values?*. arXiv. https://doi.org/10.48550/arXiv.2410.08418.

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*, 81–95. https://doi.org/10.1007/s10209-014-0348-1

Matthes, J. (2006). The need for orientation towards news media: Revising and validating a classic concept. *International Journal of Public Opinion Research, 18*(4), 422–444. https://doi.org/10.1093/ijpor/edh118

McKinney, S. (2024). Integrating artificial intelligence into citizens' assemblies: Benefits, concerns and future pathways. *Journal of Deliberative Democracy, 20*(1), 1–12. https://doi.org/10.16997/jdd.1556

Narayanan, A. (2023). *Understanding social media recommendation algorithms*. Knight First Amendment Institute at Columbia University. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

Neblo, M. A., Esterling, K. M., Kennedy, R. P., Lazer, D., & Sokhey, A. E. (2010). Who wants to deliberate—And why? *American Political Science Review, 104*(3), 566–583. https://doi.org/10.1017/S0003055410000298

Neblo, M. A., Esterling, K. M., & Lazer, D. (2018). *Politics with the people: Building a directly representative democracy*. Cambridge University Press. https://doi.org/10.1017/9781316338179

Niemeyer, S., Veri, F., Dryzek, J. S., & Bächtiger, A. (2024). How deliberation happens: Enabling deliberative reason. *American Political Science Review, 118*(1), 345–362. https://doi.org/10.1017/S0003055423000023

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819346

Oni, A. A., Oni, S., Mbarika, V., & Ayo, C. K. (2017). Empirical study of user acceptance of online political participation: Integrating civic voluntarism model and theory of reasoned action. *Government Information Quarterly, 34*(2), 317–328. https://doi.org/10.1016/j.giq.2017.02.003

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2023). What governs attitudes toward artificial intelligence adoption and governance? *Science and Public Policy, 50*(2), 161–176. https://doi.org/10.1093/scipol/scac056

Rauchfleisch, A., Jungherr, A., & Wuttke, A. (2025). Explaining public preferences for regulating artificial intelligence in election campaigns: evidence from the U.S. And Taiwan. *Telecommunications Policy*. In Press.

Rose, J., & Sæbø, Ø. (2010). Designing deliberation systems. *The Information Society, 26*(3), 228–240. https://doi.org/10.1080/01972241003712298

Sæbø, Ø., Rose, J., & Molka-Danielsen, J. (2009). eParticipation: Designing and managing political discussion forums. *Social Science Computer Review, 28*(4), 403–426. https://doi.org/10.1177/0894439309341626

Schlozman, K. L., Brady, H. E., & Verba, S. (2018). *Unequal and unrepresented: Political inequality and the people's voice in the new gilded age*. Princeton University Press. https://doi.org/10.23943/9781400890361

Siegrist, M., & Visschers, V. H. M. (2013). Acceptance of nuclear power: The Fukushima effect. *Energy Policy, 59*(August), 112–119. https://doi.org/10.1016/j.enpol.2012.07.051

Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., … Megill, C. (2023). *Opportunities and risks of LLMs for scalable deliberation with polis*. arXiv. https://doi.org/10.48550/arXiv.2306.11932.

Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics, 1*, 21–48. https://doi.org/10.1057/palgrave.cep.6110002

Steibel, F., & Estevez, E. (2015). Designing Web 2.0 tools for online public consultation. In A. Chib, J. May, & R. Barrantes (Eds.), *Impact of information society research in the global south* (pp. 243–263). Springer. https://doi.org/10.1007/978-981-287-381-1_13.

Tessler, M. H., Bakker, M. A., Jarrett, D., Sheahan, H., Chadwick, M. J., Koster, R., … Summerfield, C. (2024). AI can help humans find common ground in democratic deliberation. *Science, 386*(6719), Article eadq2852. https://doi.org/10.1126/science.adq2852

Tsai, L. L., Pentland, A., Braley, A., Chen, N., Enrıquez, J. R., & Reuel, A. (2024). Generative AI for pro-democracy platforms. In D. Huttenlocher, & A. Ozdaglar (Eds.), *An MIT exploration of generative AI: From novel chemicals to opera*. MIT. https://doi.org/10.21428/e4baedd9.5aaf489a.

Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in american politics*. Harvard University Press.

Wang, S. (2024). Public perceptions of artificial intelligence in 20 countries: Assessing individual- and country-level factors. *Cross-Cultural Research*. https://doi.org/10.1177/10693971251336803

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human?: The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Young, I. M. (2002). *Inclusion and democracy*. Oxford University Press. https://doi.org/10.1093/0198297556.001.0001

Zhang, B. (2024). Public opinion toward artificial intelligence. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford handbook of AI governance* (pp. 553–571). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.36.

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Center for the Governance of AI University of Oxford. https://doi.org/10.2139/ssrn.3312874

Andreas Jungherr holds the *Chair for Political Science, especially Digital Transformation* at the University of Bamberg and is a Director at the *Bavarian Research Institute for Digital Transformation (bidt)*. He examines the impact of digital media on politics and society, with a special focus on political communication, Artificial Intelligence, and governance. He is the author of *Retooling Politics: How Digital Media is Shaping Democracy* (with Gonzalo Rivero and Daniel Gayo-Avello, Cambridge University Press: 2020) and *Digital Transformations of the Public Arena* (with Ralph Schroeder, Cambridge University Press: 2022).

Adrian Rauchfleisch is a Professor at the Graduate Institute of Journalism, National Taiwan University. His research focuses on the interplay of politics, technology, and journalism in Asia, Europe, and the United States. His current project explores the influence of Artificial Intelligence on society across different cultural contexts.