Aligning AI with Public Values: Deliberation and Decision-Making for Multimodal LLM Governance

Anonymous ACL submission

Abstract

How AI models should deal with political topics has been discussed, but it remains challenging and requires better governance. This paper examines the governance of large language models through individual and collective deliberation, focusing on politically sensitive videos. We conducted a two-step study: interviews with 10 journalists established a baseline understanding of expert video interpretation; 114 individuals through deliberation using *Inclusive.AI*, a platform that facilitates democratic decision-making through decentralized autonomous organization (DAO) mechanisms. Our findings reveal distinct differences in interpretative priorities: while experts emphasized emotion and narrative, general public prioritized factual clarity, objectivity, and emotional neutrality. Furthermore, we examined how different governance mechanisms - quadratic vs. weighted voting and equal vs. 20/80 voting power - shape users' decision-making regarding AI behavior. Results indicate that voting methods significantly influence outcomes, with quadratic voting reinforcing perceptions of liberal democracy and political equality. Our study underscores the necessity of selecting appropriate governance mechanisms to better capture user perspectives and suggests decentralized AI governance as a potential way to facilitate broader public engagement in AI development, ensuring that varied perspectives meaningfully inform design decisions.

1 Introduction

A major criticism of AI development is the lack of transparency, particularly the insufficient documentation, and traceability in model design, specification, and deployment (Brundage et al., 2020), leading to adverse outcomes including discrimination, lack of representation, and breaches of legal regulations. Traditional social science approaches, such as interviews and surveys, often fall short in capturing user expectations due to their limitations in facilitating ongoing deliberation. Governance, in contrast, is an interdisciplinary research area that involves stakeholders, (Shneiderman, 2020; Bu et al., 2020; Rubinstein and Good, 2013; Wang et al., 2022) for structural changes, such as defining bias criteria, determining rules for dataset diversity, etc. This involves principles such as normative positions, concrete actions, and engineering practices.

AI governance literature often clusters into key themes, many borrowed from data protection and privacy fields- (1) FACT - fairness, accuracy, confidentiality, and transparency (Kemper and Kolkman, 2019; Kaminski and Malgieri, 2020; Selbst, 2021); (2) FATE - fairness, accountability, transparency, and ethics (Barocas et al., 2013); (3) privacy preservation; (4) governance, compliance, and risk (Calo, 2017; Gasser and Almeida, 2017; Scherer, 2015; Butcher and Beridze, 2019); (5) trust and safety (Biden, 2023; Shneiderman; Wang et al., 2022; Saravanakumar and Arun, 2014; Biden, 2023); and (6) alignment with human values (Ji et al., 2024; Norhashim and Hahn, 2024). Additionally, there is a growing focus on participatory AI (Young et al., 2024) leveraging existing international legal frameworks (Cihon, 2019; Maas, 2021; Wallach and Marchant, 2018; Erdélyi and Goldsmith, 2018). The AI Executive Order further highlights the need for a coordinated approach, emphasizing community engagement (Biden, 2023).

Emerging models such as Decentralized Autonomous Organizations (DAOs) (Sharma et al., 2023) also provide innovative directions for technical elements that support varied structural concepts from management science and community coordination. DAOs are blockchain-based organizations governed by smart contracts and decentralized decision-making, enabling collective governance without centralized control (Sharma et al., 2023). By leveraging transparent, automated processes with smart contract governance, DAO provides a potential empirical testbed for exploring social choice experiments in potentially improving the current AI governance structure through a computational lens (Benkler et al., 2015; Lalley and Weyl, 2018; Weyl et al., 2022; Zhang and Zhou, 2017; Weber, 2015). However, a fundamental tension exists between participatory decision-making in AI and its global, distributed nature (Young et al., 2024). DAOs present unique opportunities to address this challenge by implementing mechanisms such as social choice designs, quadratic voting, and liquid democracy (Lalley and Weyl, 2018; Weyl et al., 2022; Zhang and Zhou, 2017), while also enabling anonymous participation for diverse voices.

To examine the benefits of decentralized governance in AI development, we conduct a case study focusing on how AI systems should address politically sensitive topics. The use of LLMs in political domains has been widely debated, including their political biases (Potter et al., 2024a,b; Rozado, 2024; Feng et al., 2023; Santurkar et al., 2023). Recent studies have revealed that LLMs can influence users' political views through their interactions (Potter et al., 2024b; Fisher et al., 2024; Costello et al., 2024). While several approaches have been proposed to pursue the political neutrality of LLMs, no clear consensus has emerged (Potter et al., 2024a; Sorensen et al.); for example, many users expressed enjoyment when they are engaged in the interaction with politically leaned LLMs (Potter et al., 2024b). The conflicting views on these issues highlight the need for a deliberative process to incorporate diverse user perspectives.

This motivates our research questions: **How does the general public perceive the use of LLM in political content interpretation? How do DAO governance mechanisms influence public opinions about improving LLM design?**

We propose **Inclusive.AI**, a DAO-enabled governance, emphasizing inclusivity and human oversight in LLM design oversight. As illustrated in Figure 1, to explicitly understand users' specific expectations, the governance model allows users to deliberate on sensitive topics where LLM output can be controversial and contentious. For our experiment, we used a video from the 2020 US presidential debate as a case study to explore public preferences in governing LLM behavior (Linegar et al., 2023). To ensure secure and equitable participation, we implemented DAO infrastructure to enhance trust in the governance process. With Inclusive.AI, users first deliberate on LLM outputs, express their preferences and then participate in governance voting to guide future LLM design for political video interpretations.

Findings. Through an online experiment of 114 US internet users, our findings highlighted overlapping values between individual and collective deliberation for improving LLM output for political video content. Some factors are considered important, including, the emotions of the speaker, subjective content (e.g., who supports or opposes, composure, professionalism), and the speaker's positionality. There are some distinct differences in interpretative priorities: while experts emphasized emotion and narrative, general public prioritized factual clarity, objectivity, and emotional neutrality. Our findings also highlighted participants' perceived quality of the governance of the Inclusive.AI tool whereas voting methods significantly influence outcomes, with quadratic voting reinforcing perceptions of liberal democracy and political equality. They emphasized that quadratic voting, under equal voting power conditions, reduces the likelihood of producing unexpected outcomes compared to weighted voting. However, some were skeptical about whether the decided outcomes would be implemented in LLM models, suggesting guidelines at the government level to ensure compliance.

2 Related Work

Video Analysis in Practice & Multimodal Generative Vision Models. Videos are a rich source of information for communication (Chen and Jiang, 2019; Lin et al., 2021), driving tasks like video captioning, question answering (Yang et al., 2021), text-video retrieval (Gabeur et al., 2020; Bain et al., 2021; Anne Hendricks et al., 2017). Identifying key visual content in video-language learning remains a challenge (Buch et al., 2022; Lei et al., 2022). Political science research increasingly explores video content (Hong et al., 2021) where language models often exhibit biases in multi-modal data. Advancements in computer vision have led to foundational vision-language models, such as CLIP in numerous downstream applications, ranging from object detection to 3D applications (Bangalath et al., 2022; Liang et al., 2023; Rozenberszki et al., 2022; Ni et al., 2022), and adapted for video applications (Ni et al., 2022; Wang et al., 2021; Rasheed et al., 2023). More recently, multimodal integration has advanced with models like Flamingo (Alayrac et al., 2022), BLIP-2 (Li et al., 2023a) MiniGPT-4 (Zhu et al., 2023), and LLaVA (Liu et al., 2024) leveraging web-scale image-text

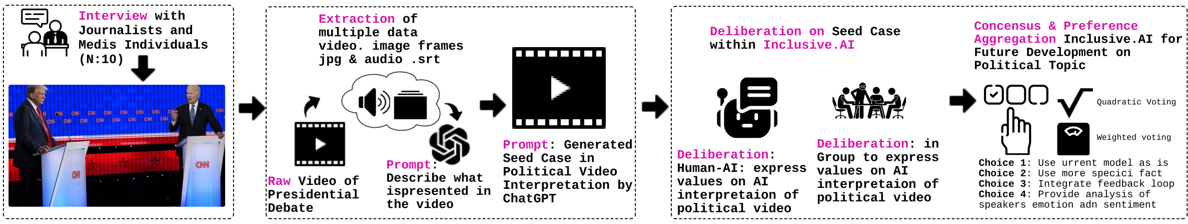


Figure 1: An overview of processes, (a) interview with experts to select suitable video example; (b) prepare seed case for experiment setup; (c) incorporate seed case into inclusive.AI system for deliberation and preference gathering (i) deliberation human-AI and group (ii) democratic voting process incorporating the voting options from deliberation of general public.

data for improved multimodal chat capabilities. Some works extend LLMs for video comprehension (Maaz et al., 2023; Radford et al., 2021; Chiang et al., 2023; Li et al., 2023b; Liu et al., 2024), introducing Video-ChatGPT, model combining a video-optimizer for enhanced understanding.

DAO as a tool for Governance and Coordination. Decentralized Autonomous Organizations (DAOs), which emerged in the mid-2010s, share commonalities with early online communities, especially those focused on open-source projects (Chohan, 2017). DAOs also draw inspiration from various models, including digital and platform cooperatives (Mannan, 2018), multi-organizational networks like keiretsus (Lincoln et al., 1996), crowdfunding platforms such as Patreon, virtual economies in games like World of Warcraft and Second Life (Lehdonvirta and Castronova, 2014), and peer-produced projects like Wikipedia (Xu and Li, 2015). DAO governance, as a human-centric digital organization, addresses key issues in social computing but can be more complex than platforms such as civic tech (Poor, 2005), and traditional online communities (Love, 2010). DAOs were designed to automate organizational processes leveraging cryptographically secured blockchain technology (Buterin, 2014). A key function of a DAO is collective decision-making - carried out through a series of proposals where members vote on organizational events using governance tokens, signifying relative influence within the DAO. Voting mechanisms like weighted and quadratic voting ensure secure, pseudonymous participation, with voters identified by on-chain addresses rather than real-world identities.

The emergence of DAOs introduces possible solutions, including classic coordination dilemmas such as preference aggregation, credible commitments, audience costs, information asymmetry, representation, and accountability (Hall and Taylor, 1996; ope, accessed on 2024). The relevance of these theories to the design of digitally-native governance institutions is a critical question (Rousseau, 1964; Dahl, 1989; Landemore, 2012). The separation of powers in DAOs helps prevent power concentration, enhance transparency, and mitigate organizational gridlock (De Montesquieu, 1989). This is increasingly relevant for AI, where inclusive decision-making is crucial throughout development lifecycle. In this work, we explore the design of DAO in AI governance for model decision-making.

3 Inclusive.AI Design and Experiment

As shown in Figure 1, our entire study includes (1) an expert interview (protocol in Appendix B) with journalists and media individuals in selecting a suitable political video [1] as a seed case for user experiments; (2) a large-scale user experiment in deliberating users' values regarding the LLM interpretation of political topics The user experiment has three main design components-(1) Human-AI interaction to deliberate on sensitive topics (e.g. presidential debate video), (2) Group discussion to engage with other to understand collective opinions (3) Governance decisions to guide future LLM model updates.

3.1 System Design

Inclusive.AI (GitHub (Anonymous, 2024)) democratic platform (Figure 2) is deployed on the Op-

[1]Since we aim to understand the general public's perception of the use of LLMs for sensitive topics, such as political content, selecting politically sensitive content for the study requires careful consideration. We leveraged experts' opinions to conform to the inclusion criteria for selecting content) by providing them with an overview of the user study goal. The inclusion criteria mentioned: (a) relevance to current events (b) Broad political video (c) contextual depth or complexity (d) authenticity of content sources. We also asked them how they would prompt the LLM tool to interpret this video. We leveraged experts' feedback to design the deliberation case.

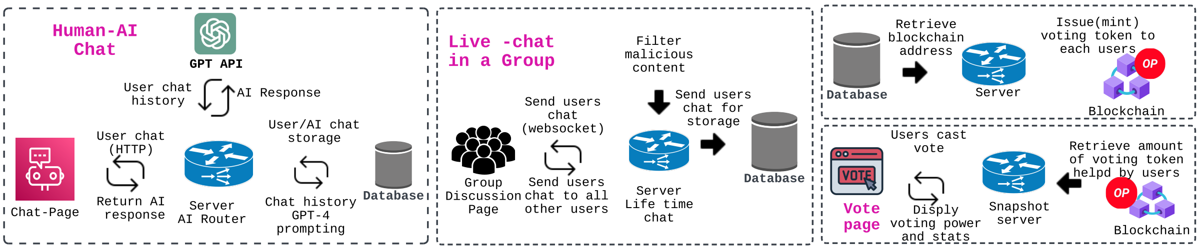


Figure 2: Incluisve.AI System Architecture

timism blockchain and integrates with a custom server, using Web3Auth (Goldreich, 1998) for authentication. Web3Auth generates a unique Multi-Party Computation (MPC) wallet for each user, derives their blockchain address, and enables message authentication for verifying participation in votes. Upon signup, they are guided to an introduction 2-minute video overview of task details and app functionality. They then proceed to Human-AI deliberation and group discussions, where a chat box with websocket connections supports real-time interactions.

For the voting page, we implemented two Vote-Token contracts using Solidity, a programming language for the Ethereum blockchain—to represent users' voting power. These tokens are *minted* to users' accounts, allowing them to vote on proposals for LLM improvement of political video. The system uses the Snapshot API to create a space for governance and ensure all the processes are transparent in Blockchain. Spaces define voting rules (e.g., duration), proposal criteria (e.g., success thresholds of proposed options to be considered for LLM improvements), and roles for admins and moderators, including who can vote or propose changes. We designed spaces for each experimental condition (each type of governance decision mechanism discussed in section 3.3). When the user allocates votes accordingly and clicks the "Cast Vote" button (in Figure 6), this triggers Web3Auth's signing library, which signs a message for Snapshot voting.

3.2 Deliberation and Decision Making

AI Guided Individual & Group Deliberation. The app begins by engaging users with an AI Value Topic related to data interpretation of a video on a political topic by GPT4 (Figure 4). This topic is based on a 6-minute clip from the 2020 US presidential debate (Anonymous, 2024) The app presented a simple question: *"Do you find the interpretation useful?" with three options (yes, no, maybe)* to stimulate further thought. Based on the user's response to the provided options, the AI continues the corresponding chat that allows users to clarify their intentions and values in natural-language conversations about AI value topics. AI resolves ambiguities through multi-turn conversations, seeking clarifications and guiding users to define their norms and expectations. Following that, users engaged in a group deliberation and learned the perspectives of others' norms (Figure 5). This group deliberation enables users to co-validate their values with a mini-public to make informed decisions. If participants are unable to introduce a topic on their own, they are encouraged to refer to the suggested topics provided by the tool. We designed the suggested topic based on the pilot experiment (in Appendix Section A)

Democratic Decision Making for Future MM-LLM Finally, users participate in a democratic decision-making process by voting (Figure 6). We designed experiments to assess varying voting methods and combinations of voting power (details in section 3.3) to examine users' perception of the quality of the process being democratic in LLM model improvement decisions. We assessed users' self-reported quality with the Variety of Democracy (V-Dem) scale (Lindberg et al., 2014). The voting was live for 48 hours.

3.3 User Experimental Design

Treatment Condition: Varying Governance Voting Design In governance decision-making, voting methods and voting power are key factors influencing outcomes, as demonstrated in DAOs and deliberative democracy (Sharma et al., 2023, 2024; Fritsch et al., 2024; Willis et al., 2022; Follesdal, 2010). To structure decision-making to aggregate people's preferences for future LLM development, we designed a 2×2 treatment condition based on two factors: voting method and voting power, each with two levels. While alternative methods like single-choice or approval voting could also be considered, it would significantly increase the number of treatment conditions and require a large participant pool to achieve statistically significant results with actionable interpretations.

More specifically, we implemented weighted voting, commonly used in DAOs (Sharma et al., 2023), where users distribute voting power across multiple options based on preference. To counterbalance traditional democratic aggregation which may disadvantage minority views, we incorporated quadratic voting - largely applied in real-world cases, such as Gitcoin's grant funding for public goods(Miller et al., 2024)–which enhances minority influence on crucial issues by allowing users to "pay" for additional votes. For instance, with quadratic voting, 4 tokens provide 2 votes, emphasizing the number of voters rather than voting power size (Lalley et al., 2016). To address voting power distribution, we compared equal distribution with a Pareto-based 20/80 split, where 20% of participants receive 80% of tokens, simulating early adopters' influence. This model reflects real-world AI deployment scenarios, where certain groups benefit disproportionately.

Thus, there were four treatment conditions- (1) Quadratic Voting token-based (Participants having the same amount of token/voting power); (2) Quadratic Voting 20% population get 80% of the token as early adopters; (3) Weighted voting Token based (participants having the same amount of token/voting power); (4) Weighted voting 20% population get 80% of the token as early adopters. The goal was to assess how these variations influence users' perceptions of the process's democratic quality and outcome.

Experimental Conditions. Participants were randomly assigned by the Inclusive.AI system to one of four governance decision-making mechanisms, forming four treatment groups. Participants didn't know the treatment group to which they had been assigned. We employed a $2 * 2$ between-subjects design with 114 participants (26-30 per condition). Participants voted on four MM-LLM update options derived from 20 pilot studies for political video interpretation: (i) keep the current model; (ii) provide more specific facts; (iii) integrate a user feedback loop; (iv) analyze speakers' emotions and sentiment (as shown in Figure 6).

3.4 Participant Demographics

We recruited participants who are USA residents. We recruited through the CloudResearch platform (Clo). This study protocol involving human subjects was approved by the Institutional Review Board (IRB). Each received $30 for their participation. We used a set of screening questions. Respondents were invited to our study if they met all three selection criteria - (1) 18 years or older; (2) country of residence USA; (3) use generative AI tool. Our study resulted in total of 114 participants (Demographics in Table 5).

4 People's Opinion on LLM Interpretation of Political Video

Journalist's Opinion of LLM in Political Video. We found several practices of journalists in interpreting political videos on their own, including- (a) fact-checking with multiple data sources and guidelines (e.g. media literacy project (lit, accessed on 2024), MSA Security (msa, accessed on 2024)), (b) involvement of expert-in-the-loop (e.g. academic scholars, senior journalists, domain experts), (c) narrative approach considered as news generation 101; (d) theoretical underpinning, such as positionality, selective exposure (Tully et al., 2022).

Experts highlighted several limitations in LLM-generated summaries of political videos, particularly the absence of human interaction cues such as tone and emotion. They noted that the lack of contextual information, including background knowledge on political debates, reduced the summary's usefulness for news content. While factually accurate, the summary failed to capture the antagonistic and dramatic dynamics of the debate, including conflicts, personal attacks, and the candidates' lack of factual references. Additionally, experts criticized its lack of storytelling and engagement, making it unsuitable for a diverse audience and insufficient in depth and impact.

General Public's Opinion. Our findings of users' interaction with the seed case on political video interpretation highlight various factors participants considered important on interpreting video content while analyzing multiple types of data (e.g. image frame, audio, etc). In group deliberation, we found that participants articulated their arguments in longer sentences, while in human-AI chats, the conversations were shorter. In individual value elicitation, we also found participants to suggest specific design recommendations of how to generate and present the LLM output rather than only pointing out what is lacking. They tend to begin their interactions with a positive tone. As the con-

Table 1: Overview of Themes of Deliberation on LLM Output of Political Video

Theme	Quote	Ind / Group
Emotion of the Speakers	"There was a heated argument in video, both speakers didn't want to give way for other to speak, Trump and moderator were talking like they were fighting, its not in the LLM output."	Indv & Group
Objectivity of The Situation	"It didn't understand situation at all, AI was superficial, capturing the scene, distinguished between speaker and their political view is important, I could have just read the subtitle instead."	Indv & Group
Desire Fact-Checking	"Fact-checking whether the debaters are saying anything of substance would greatly help in giving an accurate picture of the view, like citing some source while interpreting the video."	Indv & Group
Balance Brevity and Substance	"AI describe he video, but there was no real context, like to take away"	Group
Balancing Content & Biases	"This is one of those times when I wish AI could let itself loose just a little more, necessarily—just to the fact of acknowledging how Trump was not acting as a good steward of discussion, also the too much emphasize towards Obamacare and social support system."	Group
Organization of LLM output	"It would be easier to differentiate to have the description of two candidates side by side."	Indv
Specific Design Recommendations,	"Red highlight for content in the video that are factually wrong and green for truth."	Indv

versations progressed, participants shifted towards making recommendations and expressing concerns. In contrast, group deliberation started with a tone of concern and debate.

However, in both types of interactions, there are overlapping values emerged regarding LLM improvement for political content interpretation. This includes: the emotions of the speaker, subjective content (e.g., who supports or opposes, composure, professionalism), and the speaker's positionality. We also observed nuanced differences in individual values, for instance, participants tend to express a preference for fact-focused political LLM interpretation with specific indicators as design recommendations and emphasized the importance of clarity and organization of LLM output (Table 1 presents example quotes).

5 Experience in Democratic Governance

Preference on LLM Improvement Choices. For improving MM-LLMs in political video interpretation, participants strongly preferred *"providing more specific facts"*(choice 2), followed by *"analyzing speakers' emotions and sentiment"* (choice 4) and *"integrating a user feedback loop"* (choice 3). The consistency of choices 2 and 4 across quadratic and weighted methods indicates stable user preference (Table 2). However, in 20/80 voting power distributions, early adopters (80% power) influenced outcomes, narrowing the gap between choices 3 and 4. This suggests that in real-world governance of LLM improvements, decision-making that concentrates power among a few influential stakeholders could disproportionately shape LLM improvements, potentially misaligning with broader user preferences.

To see whether participants affiliated with different political parties had different choices and perceptions for LLM improvement on political video interpretation, we ran a linear regression controlling for voting methods. As a result, we found that, compared to Democrats, Republicans were less likely to vote for Choice 2 (i.e., provide more specific facts) with a P-value of 0.084[2] and more likely to vote for Choice 3 (i.e., integrate a user feedback loop) with a P-value of 0.054.

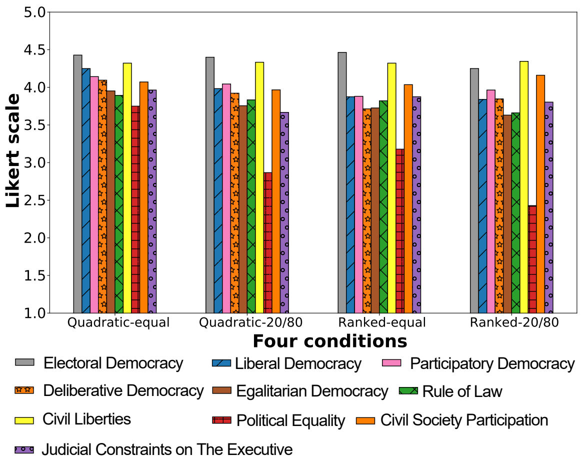


Figure 3: Users' perception of a voting mechanism (obtained through the V-Dem question lists)

General Perception of Voting Mechanism. With a 5-point Likert scale (Figure 7 in Appendix), we found participants' perceptions of the voting process usage in LLM governance where most participants were satisfied with the process regardless of voting protocols. Notably, they rated with average scores of $3.89, 4.17, 3.96,$ and 3.93 in the four voting mechanisms: quadratic+equal, quadratic+20/80, ranked+equal, and ranked+20/80. Quadratic voting and equal power distribution enhanced partic-

[2]We used the criterion of P-value< 0.1, considering the small number of participants.

Table 2: Summary stats of the ratio of tokens allocated to each voting choice (Choice 1: Keep the current model, Choice 2: Provide more specific facts, Choice 3: Integrate a user feedback loop, and Choice 4: Analyze speakers' emotions and sentiment) by users. The ratio is calculated as the percentage of tokens the user allocated to each voting option. For example, if a user allocated 20, 20, 30, 30 tokens for each voting option, the vector for the user would be (0.2,0.2,0.3,0.3).

	Choice 1		**Choice 2**		**Choice 3**		**Choice 4**	
	mean	std.	mean	std	mean	std	mean	std.
Quadratic - same (n: 29)	0.0814	0.0885	0.4300	0.3421	0.1524	0.1690	0.2597	0.2905
Quadratic -20/80 (n: 30)	0.1193	0.2197	0.3267	0.2260	0.2188	0.1956	0.3080	0.2473
Ranked - same (n: 27)	0.1193	0.1412	0.3941	0.2255	0.1896	0.1197	0.2926	0.2351
Ranked - 20/80 (n:28)	0.1395	0.1534	0.4044	0.2692	0.1877	0.1400	0.2417	0.1870

Table 3: Governance Decision-Making Experience Across Different Political Leaning

Democrats	Republicans	Independent / Unaffiliated
Progressive and Empowerment	Flexible in Distributing power	Desire for Additional Option
Ease of Use and Intuitive	New Experience and Curiosity	Ease of Use and Intuitive
Support Multiple Choices	Quantifying Perception and Thinking Critically	Weighted Voting as a Preferred Feature
Quadratic Voting Perceived as Fair	Having Influence on AI Development	Applying This Process to Other Contexts
Engaging and Enjoyable	Concerns About Complexity and Restrictions	Concerns of Fairness and Transparency
Informed and accurate decision-making	Concerns About External Influences and Bias	Concerns About the Process's Impact

ipants' trust in the decision-making process, reducing concerns about unexpected outcomes. As participants shared *"I split my votes across multiple issues, but I think this is the purpose—to vote carefully for the option I care about most. It allows stronger opinions on some issues. the square thing I like, so even if sometimes someone had more token than me, that's actually not the number that would apply rather square root."* A linear regression analysis confirmed this effect: the coefficient for quadratic $0.4772 (P = 0.013)$, for same was $0.4002 (P = 0.038)$. The linear regression considering the interaction also demonstrated statistical significance; the coefficient of *quadratic*×*same* was 1.1548 with a P-value of 0.002.

Quality of Decision-Making Process of Different Democracies. We examined participants' perceptions of LLM governance using the Varieties of Democracy (V-Dem) (Figure 3). Quadratic voting significantly enhanced perceptions of liberal democracy (coefficient$= 0.2549$, P-value$= 0.036$) and political equality (coefficient$= 0.4895$, P-value$= 0.037$). As noted, *"The voting was inclusive—I would like this process in chatGPT like system where they broadcast such voting time to time to get some signal from users rather deploying by themselves only."* This supports the argument that active user participation in AI decision-making can enhance legitimacy, rather than centralized deployment by developers. Voting power distribution further reinforced perceptions of political equality (coefficient$= 0.8091$, P-value< 0.001), with linear regression considering interaction confirming its significance (coefficient of *same*$= 0.7500$, P-value$= 0.019$). This highlights the need of fair representation in AI oversight, where users regardless of their expertise or influence should have a say in shaping AI behavior.

Relationship Between Users' Value Towards AI and Their Perceived Democracy Value Participants who found LLM personally relevant were more likely to view the DAO-enabled voting process as highly participatory (Figure 8)(Pearson Corr$= -0.4426$, P-value< 0.001). This underscores the need for AI systems to establish personal relevance with users, potentially through more user-centered political content moderation. Perceptions of deliberative democracy were strongly linked to trust in AI companies(Pearson Corr$= 0.4422$, P-value< 0.001) and perceived AI risks (Pearson Corr$= 0.5142$, P-value< 0.001). This suggests that skepticism about AI risks coexists with the belief that AI governance should involve ongoing public discourse. For LLM governance, this emphasizes the need for mechanisms that allow users to contest, audit, and deliberate on AI-generated political content, rather than simply consuming it. Participants who valued civil liberties also emphasized the importance of diverse datasets in AI training (Pearson Corr$= 0.4646$, P-value< 0.001), uncertainty handling by AI developers (Pearson Corr$= 0.5326$, P-value< 0.001), the perceived AI risks (Pearson Corr$= 0.4407$, P-value< 0.001), and the desired reliance on AI (Pearson Corr$= 0.4950$, P-value< 0.001). This underscores the necessity of dataset diversity, bias mitigation, and AI uncertainty management in political content generation.

Attitude Towards Voting Mechanisms. Participants had key attitude in applying voting mechanisms to MM-LLM governance, including (1) progressive and fair process, (2) methods to show the strength of preference, (3) support multiple choices with a unique voice, (4) quantifying perception, (5) Inclusive.AI as practical AI governance applications (e.g., aligning with public preferences). We also found differences in the perception among political parties. Republicans tended to feel significantly more that they could contribute to shaping the space of generative AI models through this process when compared to Democrats (linear coefficient= 0.521, P-value= 0.007). Qualitative analysis of survey data also revealed some differing perspectives (Figure 3). Democrats emphasized empowerment, ease of use, and engagement emphasizing positive experience. In contrast, republicans prioritized functional and individual priorities like flexible voting power designs, on option to quantify perception through voting, and curiosity. Republican and independent participants also raised concerns about complexity, external influences (majority bias as a good way to go), and post-vote transparency regarding AI developers' implementation of decisions. However, these findings are not indicative of broader political divisions due to the low frequency of such experiences.

6 Discussion & Conclusion

Our findings underscore two key recommendations for practitioners aligning with LLMs and how to engage users in governance.

DAO as a Technical LLM Governance Solution. Transparency in LLM design decisions is utmost importance for aligning AI systems with societal expectations (Mitchell et al., 2019; Liesenfeld et al., 2023). To do that, it's crucial not only to gain a deeper understanding of public perceptions regarding AI but also to devise methods that actively involve the community in the decision-making processes governing AI technologies. Inclusive.AI tool, underpinned by the DAO mechanism, offers an avenue to actively involve people in governing llm with empirical evidence while presented with a sensitive topic like political video. DAO mechanisms, as digital-first entities, employ mechanisms like initiating proposals, nuanced voting methods, and blockchain-based coordination (Sharma et al., 2023), offering a structured approach to AI governance (Koster et al., 2022), a concept endorsed by industry leaders such as OpenAI, Meta, and federal agencies (ope, accessed on 2024; Biden, 2023).

A standout feature was our system's Voting method, in which participants found effectively representing their voice directly impacting AI model decisions for future improvements (Arts and Tatenhove, 2004). Participants recognized the potential of these methods in helping developers and government bodies align more closely with public preferences. However, skepticism remains about whether their votes would translate into real changes in AI models. This highlights the need for government-level guidelines to ensure system compliance and evidence through future audits.

Continuous Human Involvement for LLM Model Adaptability. Our research, drawing on insights from experts in news production reveals that video analysis in media coverage remains largely reliant on manual processes and human intervention. Critical frameworks like positionality (Callison and Young, 2019) and selective exposure balance are essential for ensuring accurate and contextually rich video interpretation, particularly in political reporting. Experts emphasize the need for diverse perspectives and contextual depth to prevent biases and ensure political content reflects a broad spectrum of viewpoints (Blumler and Kavanagh, 1999; Jacobs and Townsley, 2011)

Our findings from public deliberation and LLM governance decision-making illustrate how political affiliation shapes perceptions of LLM-generated political content. For instance, Republicans were less likely than Democrats to vote for providing more specific facts and instead favored integrating a user feedback loop for LLM improvement. InclusiveAI platform facilitate on imitating natural human interaction among people acknowledging that conflicting interests and preferences. Rather than seeking consensus on the topic, participants engaged in discussions that helped them identify compromises and make informed voting decisions. We suggest that inclusive AI systems could be integrated into LLM tools, such as ChatGPT, allowing users and experts to propose real-time adjustments and engage with broader communities when necessary. It highlights a potential future where people continuously engage in shaping functionality of AI systems with evolving needs. Potential risks of this research include political bias reinforcement due to differing perceptions of LLM-generated content, potentially deepening ideological divides.

Limitations

Our study has several limitations. Since participants were recruited from Cloudresearch and may not have been available at the same time. The asynchronous 48-hour participation window may have disrupted discussion flow and reduced engagement, even with a group chatbox supporting both synchronous and asynchronous communication.

In our study, we designed the proposals with voting options for LLM improvement for political topic derived from pilot studies. Predefining voting options which may limit the dynamic needs of the participants. Future designs could support broader participation by enabling AI-mediated, real-time generation of voting options during ongoing deliberations. Some participants struggled with terms like quadratic voting, which affected their decisions despite explanatory materials.

Ethical Considerations

This This study protocol involving human subjects was approved by the Institutional Review Board (IRB). The data collection and transcription generation was anonymous to preserve privacy of the users. This study explores decentralized governance mechanisms in decision-making for LLM improvement by engaging users, particularly in politically sensitive contexts. The InclusiveAI tool with transparent design, equitable participation can allow to shape AI with broader perspectives. This also has a future potential to potentially involve regulatory oversight for for the responsible implementation.

References

Cloudresearch. https://www.cloudresearch.com/. [Accessed 15-02-2025].

accessed on 2024. Democratic input to ai.

accessed on 2024. Msa security.

accessed on 2024. News literacy project.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. 2022. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736.

Lisa Anne Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell, and Bryan Russell. 2017. Localizing moments in video with natural language. In *Proceedings of the IEEE international conference on computer vision*, pages 5803–5812.

Anonymous. 2024. GitHub. https://github.com/AccountProject/Inclusive.AI-MM_LLM. [Accessed 14-02-2025].

Bas Arts and Jan Van Tatenhove. 2004. Policy and power: A conceptual framework between the 'old'and 'new'policy idioms. *Policy sciences*, 37:339–356.

Max Bain, Arsha Nagrani, Gül Varol, and Andrew Zisserman. 2021. Frozen in time: A joint video and image encoder for end-to-end retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1728–1738.

Hanoona Bangalath, Muhammad Maaz, Muhammad Uzair Khattak, Salman H Khan, and Fahad Shahbaz Khan. 2022. Bridging the gap between object and image-level representations for open-vocabulary detection. *Advances in Neural Information Processing Systems*, 35:33781–33794.

Solon Barocas, Sophie Hood, and Malte Ziewitz. 2013. Governing algorithms: A provocation piece. *Available at SSRN 2245322*.

Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. 2015. Peer production: A form of collective intelligence. *Handbook of collective intelligence*, 175.

Joseph R Biden. 2023. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence.

Jay G Blumler and Dennis Kavanagh. 1999. The third age of political communication: Influences and features. *Political communication*, 16(3):209–230.

Miles Brundage, Shahar Avin, Jasmine Wang, Haydn Belfield, Gretchen Krueger, Gillian Hadfield, Heidy Khlaaf, Jingying Yang, Helen Toner, Ruth Fong, et al. 2020. Toward trustworthy ai development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.

Fei Bu, Nengmin Wang, Bin Jiang, and Huigang Liang. 2020. "privacy by design" implementation: Information system engineers' perspective. *International Journal of Information Management*, 53:102124.

Shyamal Buch, Cristóbal Eyzaguirre, Adrien Gaidon, Jiajun Wu, Li Fei-Fei, and Juan Carlos Niebles. 2022. Revisiting the" video" in video-language understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2917–2927.

James Butcher and Irakli Beridze. 2019. What is the state of artificial intelligence governance globally? *The RUSI Journal*, 164(5-6):88–96.

Vitalik Buterin. 2014. Daos, dacs, das and more: An incomplete terminology guide. *Ethereum Blog*, 6:2014.

Candis Callison and Mary Lynn Young. 2019. *Reckoning: Journalism's limits and possibilities*. Oxford University Press.

Ryan Calo. 2017. Artificial intelligence policy: a primer and roadmap. *UCDL Rev.*, 51:399.

Shaoxiang Chen and Yu-Gang Jiang. 2019. Motion guided spatial attention for video captioning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8191–8198.

Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2(3):6.

Usman W Chohan. 2017. The decentralized autonomous organization and governance issues. *Available at SSRN 3082055*.

Peter Cihon. 2019. Standards for ai governance: international standards to enable global coordination in ai research & development. *Future of Humanity Institute. University of Oxford*, pages 340–342.

Thomas H Costello, Gordon Pennycook, and David G Rand. 2024. Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714):eadq1814.

Robert Dahl. 1989. Democracy and its critics yale university press. *New Haven & London*.

Charles De Montesquieu. 1989. *Montesquieu: The spirit of the laws*. Cambridge University Press.

Olivia J Erdélyi and Judy Goldsmith. 2018. Regulating artificial intelligence: Proposal for a global solution. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 95–101.

Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. *arXiv preprint arXiv:2305.08283*.

Jillian Fisher, Shangbin Feng, Robert Aron, Thomas Richardson, Yejin Choi, Daniel W Fisher, Jennifer Pan, Yulia Tsvetkov, and Katharina Reinecke. 2024. Biased AI can Influence Political Decision-Making. *arXiv preprint arXiv:2410.06415*.

Andreas Follesdal. 2010. The place of self-interest and the role of power in the deliberative democracy. *Journal of political philosophy*, 18(1):64–100.

Robin Fritsch, Marino Müller, and Roger Wattenhofer. 2024. Analyzing voting power in decentralized governance: Who controls daos? *Blockchain: Research and Applications*, page 100208.

Valentin Gabeur, Chen Sun, Karteek Alahari, and Cordelia Schmid. 2020. Multi-modal transformer for video retrieval. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pages 214–229. Springer.

Urs Gasser and Virgilio AF Almeida. 2017. A layered model for ai governance. *IEEE Internet Computing*, 21(6):58–62.

Oded Goldreich. 1998. Secure multi-party computation. *Manuscript. Preliminary version*, 78(110):1–108.

Peter A Hall and Rosemary CR Taylor. 1996. Political science and the three new institutionalisms. *Political studies*, 44(5):936–957.

James Hong, Will Crichton, Haotian Zhang, Daniel Y Fu, Jacob Ritchie, Jeremy Barenholtz, Ben Hannel, Xinwei Yao, Michaela Murray, Geraldine Moriba, et al. 2021. Analysis of faces in a decade of us cable tv news. In *KDD'21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*.

Ronald N Jacobs and Eleanor Townsley. 2011. *The space of opinion: Media intellectuals and the public sphere*. Oxford University Press.

Jiaming Ji, Mickel Liu, Josef Dai, Xuehai Pan, Chi Zhang, Ce Bian, Boyuan Chen, Ruiyang Sun, Yizhou Wang, and Yaodong Yang. 2024. Beavertails: Towards improved safety alignment of llm via a human-preference dataset. *Advances in Neural Information Processing Systems*, 36.

Margot E Kaminski and Gianclaudio Malgieri. 2020. Multi-layered explanations from algorithmic impact assessments in the gdpr. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 68–79.

Jakko Kemper and Daan Kolkman. 2019. Transparent to whom? no algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14):2081–2096.

Raphael Koster, Jan Balaguer, Andrea Tacchetti, Ari Weinstein, Tina Zhu, Oliver Hauser, Duncan Williams, Lucy Campbell-Gillingham, Phoebe Thacker, Matthew Botvinick, and Christopher Summerfield. 2022. Human-centred mechanism design with Democratic AI. 6(10):1398–1407.

Steven P Lalley and E Glen Weyl. 2018. Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*, volume 108, pages 33–37. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Steven P Lalley, E Glen Weyl, et al. 2016. Quadratic voting. *Available at SSRN*.

Hélène Landemore. 2012. *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.

Vili Lehdonvirta and Edward Castronova. 2014. *Virtual economies: Design and analysis*. Mit Press.

Jie Lei, Tamara L Berg, and Mohit Bansal. 2022. Revealing single frame bias for video-and-language learning. *arXiv preprint arXiv:2206.03428*.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023a. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR.

KunChang Li, Yinan He, Yi Wang, Yizhuo Li, Wenhai Wang, Ping Luo, Yali Wang, Limin Wang, and Yu Qiao. 2023b. Videochat: Chat-centric video understanding. *arXiv preprint arXiv:2305.06355*.

Feng Liang, Bichen Wu, Xiaoliang Dai, Kunpeng Li, Yinan Zhao, Hang Zhang, Peizhao Zhang, Peter Vajda, and Diana Marculescu. 2023. Open-vocabulary semantic segmentation with mask-adapted clip. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7061–7070.

Andreas Liesenfeld, Alianda Lopez, and Mark Dingemanse. 2023. Opening up chatgpt: Tracking openness, transparency, and accountability in instruction-tuned text generators. In *Proceedings of the 5th international conference on conversational user interfaces*, pages 1–6.

Xudong Lin, Gedas Bertasius, Jue Wang, Shih-Fu Chang, Devi Parikh, and Lorenzo Torresani. 2021. Vx2text: End-to-end learning of video-based text generation from multimodal inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7005–7015.

James R Lincoln, Michael L Gerlach, and Christina L Ahmadjian. 1996. Keiretsu networks and corporate performance in japan. *American sociological review*, pages 67–88.

Staffan I Lindberg, Michael Coppedge, John Gerring, and Jan Teorell. 2014. V-dem: A new way to measure democracy. *Journal of Democracy*, 25(3):159–169.

Mitchell Linegar, Rafal Kocielnik, and R Michael Alvarez. 2023. Large language models and political science. *Frontiers in Political Science*, 5:1257092.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024. Visual instruction tuning. *Advances in neural information processing systems*, 36.

Robert Love. 2010. *Linux kernel development*. Pearson Education.

Matthijs M Maas. 2021. Aligning ai regulation to sociotechnical change. *Oxford Handbook on AI Governance (Oxford University Press, 2022 forthcoming)*.

Muhammad Maaz, Hanoona Rasheed, Salman Khan, and Fahad Shahbaz Khan. 2023. Video-chatgpt: Towards detailed video understanding via large vision and language models. *arXiv preprint arXiv:2306.05424*.

Morshed Mannan. 2018. Fostering worker cooperatives with blockchain technology: Lessons from the colony project. *Erasmus L. Rev.*, 11:190.

Joel Miller, Chris Kanich, and E Glen Weyl. 2024. A case study in plural governance design. In *appear at the Pluralistic Alignment Workshop at NeurIPS*.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229.

Bolin Ni, Houwen Peng, Minghao Chen, Songyang Zhang, Gaofeng Meng, Jianlong Fu, Shiming Xiang, and Haibin Ling. 2022. Expanding language-image pretrained models for general video recognition. In *European Conference on Computer Vision*, pages 1–18. Springer.

Hakim Norhashim and Jungpil Hahn. 2024. Measuring human-ai value alignment in large language models. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1063–1073.

Nathaniel Poor. 2005. Mechanisms of an online public sphere: The website slashdot. *Journal of computer-mediated communication*, 10(2):JCMC1028.

Yujin Potter, Yejin Choi, David Rand, and Dawn Song. 2024a. LLMs' Potential Influences on Our Democracy: Challenges and Opportunities. In *ICLR Blogposts 2025*. Accessed: 2025-01-02.

Yujin Potter, Shiyang Lai, Junsol Kim, James Evans, and Dawn Song. 2024b. Hidden Persuaders: LLMs' Political Leaning and Their Influence on Voters. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Hanoona Rasheed, Muhammad Uzair Khattak, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. 2023. Fine-tuned clip models are efficient video learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6545–6554.

Jean-Jacques Rousseau. 1964. The social contract (1762). *Londres*.

David Rozado. 2024. The political preferences of LLMs. *arXiv preprint arXiv:2402.01789*.

David Rozenberszki, Or Litany, and Angela Dai. 2022. Language-grounded indoor 3d semantic segmentation in the wild. In *European Conference on Computer Vision*, pages 125–141. Springer.

Ira S Rubinstein and Nathaniel Good. 2013. Privacy by design: A counterfactual analysis of google and facebook privacy incidents. *Berkeley Tech. LJ*, 28:1333.

Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? In *International Conference on Machine Learning*, pages 29971–30004. PMLR.

C Saravanakumar and C Arun. 2014. Survey on interoperability, security, trust, privacy standardization of cloud computing. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pages 977–982. IEEE.

Matthew U Scherer. 2015. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, 29:353.

Andrew D Selbst. 2021. An institutional view of algorithmic impact assessments. *Harv. JL & Tech.*, 35:117.

Tanusree Sharma, Yujin Kwon, Kornrapat Pongmala, Henry Wang, Andrew Miller, Dawn Song, and Yang Wang. 2023. Unpacking how decentralized autonomous organizations (daos) work in practice. *arXiv preprint arXiv:2304.09822*.

Tanusree Sharma, Yujin Potter, Kornrapat Pongmala, Henry Wang, Andrew Miller, Dawn Song, and Yang Wang. 2024. Future of algorithmic organization: Large-scale analysis of decentralized autonomous organizations (daos). *arXiv preprint arXiv:2410.13095*.

Ben Shneiderman. 113th note on human-centered ai.

Ben Shneiderman. 2020. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 10(4):1–31.

Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell L Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, et al. Position: A Roadmap to Pluralistic Alignment. In *Forty-first International Conference on Machine Learning*.

Melissa Tully, Adam Maksl, Seth Ashley, Emily K Vraga, and Stephanie Craft. 2022. Defining and conceptualizing news literacy. *Journalism*, 23(8):1589–1606.

Wendell Wallach and Gary E Marchant. 2018. An agile ethical/legal model for the international and national governance of ai and robotics. *Association for the Advancement of Artificial Intelligence*.

Mengmeng Wang, Jiazheng Xing, and Yong Liu. 2021. Actionclip: A new paradigm for video action recognition. *arXiv preprint arXiv:2109.08472*.

Tian Wang, Carol Mullins Hayes, and Masooda Bashir. 2022. Developing a framework of comprehensive criteria for privacy protections. In *Future of Information and Communication Conference*, pages 905–918. Springer.

Rolf H Weber. 2015. Realizing a new global cyberspace framework. *Normative Foundations and Guiding Principles*.

E Glen Weyl, Puja Ohlhaver, and Vitalik Buterin. 2022. Decentralized society: Finding web3's soul. *Available at SSRN 4105763*.

Rebecca Willis, Nicole Curato, and Graham Smith. 2022. Deliberative democracy and the climate crisis. *Wiley Interdisciplinary Reviews: Climate Change*, 13(2):e759.

Bo Xu and Dahui Li. 2015. An empirical study of the motivations for content contribution and community participation in wikipedia. *Information & management*, 52(3):275–286.

Antoine Yang, Antoine Miech, Josef Sivic, Ivan Laptev, and Cordelia Schmid. 2021. Just ask: Learning to answer questions from millions of narrated videos. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1686–1697.

Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. Participation versus scale: Tensions in the practical demands on participatory ai. *First Monday*.

Bingsheng Zhang and Hong-Sheng Zhou. 2017. Brief announcement: Statement voting and liquid democracy. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 359–361.

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.

A Additional Results

Pilot Experiment. We conducted a pilot study using the "Inclusive.AI" tool to facilitate deliberation on AI-related value topics presented in video format. For this pilot, the topic was political, using a clip from the 2020 U.S. presidential debate. The tool guided users through a process of eliciting their values and expectations for LLM outputs on sensitive topics via individual and group deliberation. It also provided a platform for users to share their preferences on how MM-LLMs should function in the future, as detailed in Section 3.2.

The pilot study provided valuable insights to improve the tool. For example, the group deliberation feature initially didn't include suggested topics, which led to difficulties in starting discussions. Participants recommended including suggested topics in the group-live chat feature, leading us to a design update in the Inclusive.AI tool for the main experiment. The pilot also helped refine the MM-LLM update options used to gather participants' preferences in democratic decision-making. The initial options, based on the literature, included: (a) Use the current model as is, (b)Context-Aware Adaptation, (c)Use Feedback Loop Integration, and (d)Advanced Modality Integration Technique. Feedback from the pilot deliberation on the political video topic led to revisions for the main study, resulting in the following refined options: (a)Use the current model as is, (b)Provide more specific facts," (c)Integrate user feedback loop," and (d)Provide analysis of the speaker's emotion and sentiment. We performed a thematic analysis of the pilot data to identify key themes in the deliberations.

Recruitment and Experts Background We interviewed media scholars and journalists as our expert reference group due to their experience with various data types, including text, images, and videos, particularly analyzing complex and sensitive topics, like US presidential debates. We recruited 10 US-based experts through personal connections and word of mouth. Experts in this study come from diverse backgrounds in journalism, media, and communication, with an equal split between males and females. Among them are Ph.D researchers specializing in media studies in urban design, and political economy, journalist focused on video media who had experience in the 2020 election coverage; medical misinformation within local communities and journalists and videographers who have covered Tesla, police issues, and local TV media, offering a unique blend of skills and perspectives.

Details: Experts' Interview Our two primary objectives of the experts' interview were: (i) to have a baseline of how experts envision the use of MM-LLMs for interpreting political videos to the general public and (ii) to incorporate expert feedback into the development of our methodological approach, including criteria for selecting video examples for the study. Each interview took around 1 hour.

In the first set of questions, we inquired about their primary expertise and experience with various data types, including video. This helped us understand their approach to handling different media, covering real-time events, managing diverse data for tasks like media report writing, and the factors that influence the quality of their reporting. In the next set of questions, we showed them a political video of US presidential debate and asked them to interpret it using multimodal data (e.g., audio, visuals, closed captions). We asked, *"Can you walk me through the process you employ to analyze the video content to write a report?"* Following this, we asked their opinion on using LLMs for video analysis We then showed them how LLMs (ChatGPT) interpreted the same video and asked for their thoughts to identify the benefits, limitations, and critical factors in interpreting contentious topics.

Since we aim to understand the general public's perception of the use of MM-LLMs in better designing models for sensitive topics, such as political content, selecting politically sensitive content for the study requires careful consideration. We leveraged experts' opinions to conform to the inclusion criteria for selecting content (details in section **??**) by providing them with an overview of the user study goal. We also asked them how they would prompt the LLM tool to interpret this video. We leveraged experts' feedback to design the deliberation case (details in section 3.2).

How Experts Would Prompt to Analyze the Video? Experts suggested various ways they would prompt ChatGPT to analyze a presidential debate. Most would start with a general question like, *Can you help me to summarize what they are talking about?"* E1 mentioned that she would first ask for a summary and then follow up with, *"If I get the output, I might ask something else."* Similarly, E2 would prompt, *"Provide a summary of the videos and the point of each person in this content."* Two experts, like E3, preferred more detailed instructions, saying, *'Give me a short news brief about the presidential debate between Trump and Biden about health care policy and Obamacare. Also, capture some of the tough visual aspects, describe some of the back-and-forth banter between the moderator and Trump, and the personal attacks where people can't get a word in."* E6 would ask meta-questions to utilize ChatGPT in a journalism context: *"I normally wouldn't use ChatGPT to analyze only one video unless I have a hundred. I would want to see patterns across videos. I would ask it to analyze how many times there are interruptions, how long candidates talk over each other, and other specific metrics."*

B Experts' Interview

In this section, we present the questions that we asked during the expert interview.

B.1 General Introduction

1. Could you briefly talk about your primary area of expertise in communication, journalism, or media studies?

2. What kind of media do you usually work on? Do you ever work on video content? Can you share a recent experience with video content and describe what it was about?

3. How do you choose videos for your work? (This question is based on an earlier response about the type of work they do with video.)

B.2 Assessment of Videos by Communication, Media Scholars, or Journalists

1. When reviewing events (e.g., live or video) related to complex subjects such as politics, what key factors do you consider in drafting/generating an article on this event?

2. What does "good video analysis" mean to you?

3. Can you walk me through the process you employ to analyze video content to write a report?

4. Please consider this video (a provided video). Feel free to analyze it manually or with any existing tool you typically use.

5. Could you please write your interpretation of this video content and share it afterward?

B.3 Perceptions of Large Language Models (LLMs) for Video Analysis and Assessment of LLM-Generated Video Analysis

1. How familiar are you with the use of large language models (LLMs), such as ChatGPT, for video analysis?

2. What do you think about the idea of using LLMs for video analysis? What are the pros and cons in your opinion?

3. **Demo:** Show a sample video analysis generated by ChatGPT.

4. Now read this analysis result from the LLM for the same video. What do you think about this analysis?

5. If the analysis did not match your expectations, why?

6. If the analysis met your expectations, why?

7. Based on your review of the video and the LLM-generated response, what criteria do you consider necessary for a good video analysis result?

8. If you were to use ChatGPT for video analysis, how would you prompt it?

B.4 Use Cases of AI in Communication and Journalism

1. Can you share any current use cases where AI has been effectively integrated into communication or journalism practices?

2. How do you see the role of AI evolving in the field of journalism and media studies over the next five years?

C Survey Study Protocol

In this section, we present the survey questions used in the Inclusive.AI study, which involved 114 participants.

Governance Survey Questions We'd like to understand your voting experience. Below, we'll present a series of statements related to your voting experience and different voting methods and voting power you have used. Please indicate your level of agreement using the Likert scale provided.

Please use the following scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree.

- The decision-making process was indecisive.

- The decision-making process was good at maintaining order.

- The decision-making process may have problems, but it's better than any other form of government.

Please rate your attitude toward governance components such as voting methods (e.g., quadratic, ranking), voting power, etc.

- I found the voting method (Weighted ranking/Quadratic) meaningful to include my voice.
- I felt that I could contribute to shaping the space of the Generative AI model.
- I found this voting method relevant to the purpose of the proposal or proposal type.
- I found this voting power/token distribution (e.g., equal power, variable power) meaningful in including my voice.
- I found the voting method (Weighted ranking/Quadratic) fair.
- I felt I have some power to affect change in Generative AI future development.
- I found voting power distribution among users equitable.
- I felt the voting power distribution can result in unexpected outcomes.

Open-Ended Questions

- Please explain how you found the voting process to share your preferences on the video analysis by ChatGPT.
- What do you think is the impact of your contributions on designing a Generative AI Model that reflects informed public consensus?
- What are the potential benefits of personalizing generated video analysis by ChatGPT to align with your preferences?
- What are your concerns, if any, about analyzing video by ChatGPT?

Democratic Decision-Making Measures
Electoral Democracy:

- I believe that the voting process was free and fair.
- I felt all users had the right to vote.

Liberal Democracy:

- I believe AI models will operate independently without interference from the development team.
- I felt free to provide feedback on the AI model update without fear of repercussions.

Participatory Democracy:

- I felt that I had ample opportunities to influence the AI model update process beyond just voting.
- I felt that my feedback matters in the decisions made for AI model updates.
- I believe AI model update decisions will reflect the needs and preferences of the user community.

Deliberative Democracy:

- AI model update decisions are made after thorough discussion with the user community.
- There is a culture of open dialogue and discussion in the AI model update community.
- I believe developers of this AI model will prioritize user interests over their own preferences.

Egalitarian Democracy:

- I felt, regardless of my background, I have equal influence in the AI model update decision process.
- I felt large corporations or specific user groups do not have undue influence over AI model update decision processes.

Rule of Law:

- I believe developers will be held accountable for flaws or biases in the AI model updates after this decision process.
- The decision process treats every user's input equally, regardless of their status.

Civil Liberties:

- I felt free to express my opinions on AI model updates without fear.
- I felt free to participate in any community or forum discussing the AI model update decision process.

Political Equality:

- Wealthy individuals do not have more political influence than ordinary citizens.

- All ethnic and religious groups have equal political rights and influence.

Civil Society Participation:

- User communities play an active role in shaping AI model update policies.
- The development team actively seeks input from user groups and communities.

Political Ideology:

- What are the three political issues that matter to you?
- On a scale of 1 (strongly disagree) to 5 (strongly agree), please rate your satisfaction with the current political climate.
- How important is politics in your daily life? (Options: Very important, Somewhat important, Neutral, Not very important, Not at all important)
- Which political party do you most identify with? (Options: Republican, Democratic, Independent, Libertarian, Green, Other)
- How would you describe your political orientation? (Options: Very conservative, Somewhat conservative, Moderate, Somewhat liberal, Very liberal, Not sure, Prefer not to say)

Demographic Questions:

- What is your age range?
- What is your gender identity?
- Are you currently enrolled in an educational institution?
- What is your highest level of education completed?
- Please select your racial or ethnic background.
- How frequently do you use technology or digital devices?
- How often do you use an AI assistant such as ChatGPT?

AI Value Questions

Likert Scale on AI Representation and Customization:

- AI models should prioritize generating diverse outputs to represent a wide range of individuals.
- Customization options, like specifying gender or ethnicity, are vital for inclusive AI-generated videos.
- A diverse dataset in AI training is essential to prevent bias and ensure fair representation when analyzing videos on political topics.
- AI developers should prioritize uncertainty handling to avoid assumptions and ensure diverse outputs.

Trust and Personalization of Generative AI:

- The use case is not relevant to me.
- I feel AI could infringe on my representation.
- I do not fully trust the abilities of an AI model.
- The use case is too important to let the AI model decide for me.
- I am concerned that it would not be exactly clear how video analyses are produced by AI.
- I believe that AI in general would treat me fairly when making decisions and suggestions.
- If I have any problem with AI decisions, I believe OpenAI would take necessary measures.

C.1 Figures and Tables

Table 4: Experts demographics and background.

ID	*Gender*	*Age*	*Media Background*
E1	Female	25-34	TV News, Police issues
E2	Female	25-34	Environment, Architecture
E3	Female	25-34	Local, Under-represented
E4	Male	35-44	Political, Election
E5	Female	35-44	Weather, Political
E6	Male	35-44	TV Media
E7	Female	25-34	Economy, Tesla
E8	Male	45-54	Public Communication
E9	Male	25-34	Political
E10	Male	25-34	Local, Urban Design

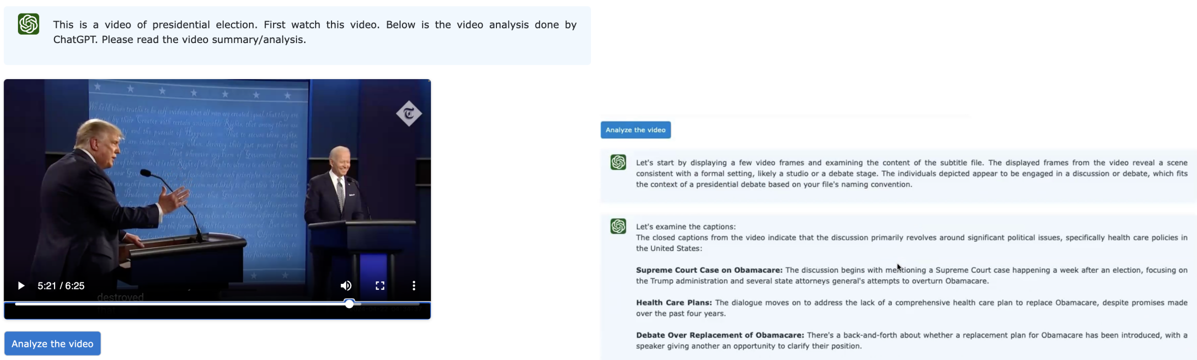
Table 5: Participants' demographics ($n = 114$)

Gender (%)			**Age (%)**				**Race (%)**				
Woman	Man	Non-binary	18-24	25-34	35-44	45-54	White	Black	Asian	Latin	Others
45.6	52.6	1.8	21.1	39.5	27.2	12.3	52.6	12.3	21.9	10.5	2.63

Education (%)					
High school	Bachelor	Masters/professional	Doctorate	College/vocational training	Others
14.0	41.2	12.3	2.6	26.3	3.5

Political Orientation (%)				
Very conservative	Conservative	Moderate	Liberal	Very liberal
5.3	21.1	22.8	35.1	15.8

Political Party (%)			
Republic party	Democratic party	Libertarian party	Independent/Unaffiliated
21.9	50.9	2.6	24.6



(a) AI value topic in video format where value topic is political- a clip from US presidential debate 2020

(b) Generated response of ChatGPT based on the video with multiple data (e.g. video frames, audio, closed caption)

Figure 4: Users' workflow in Human-AI interaction on AI value topic

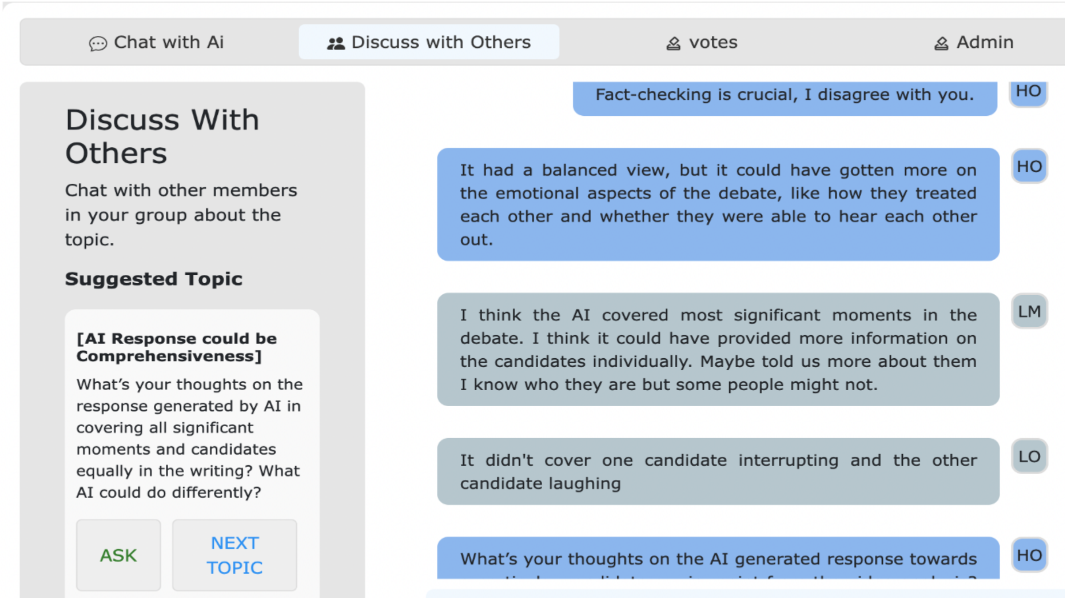


Figure 5: Discussion interface

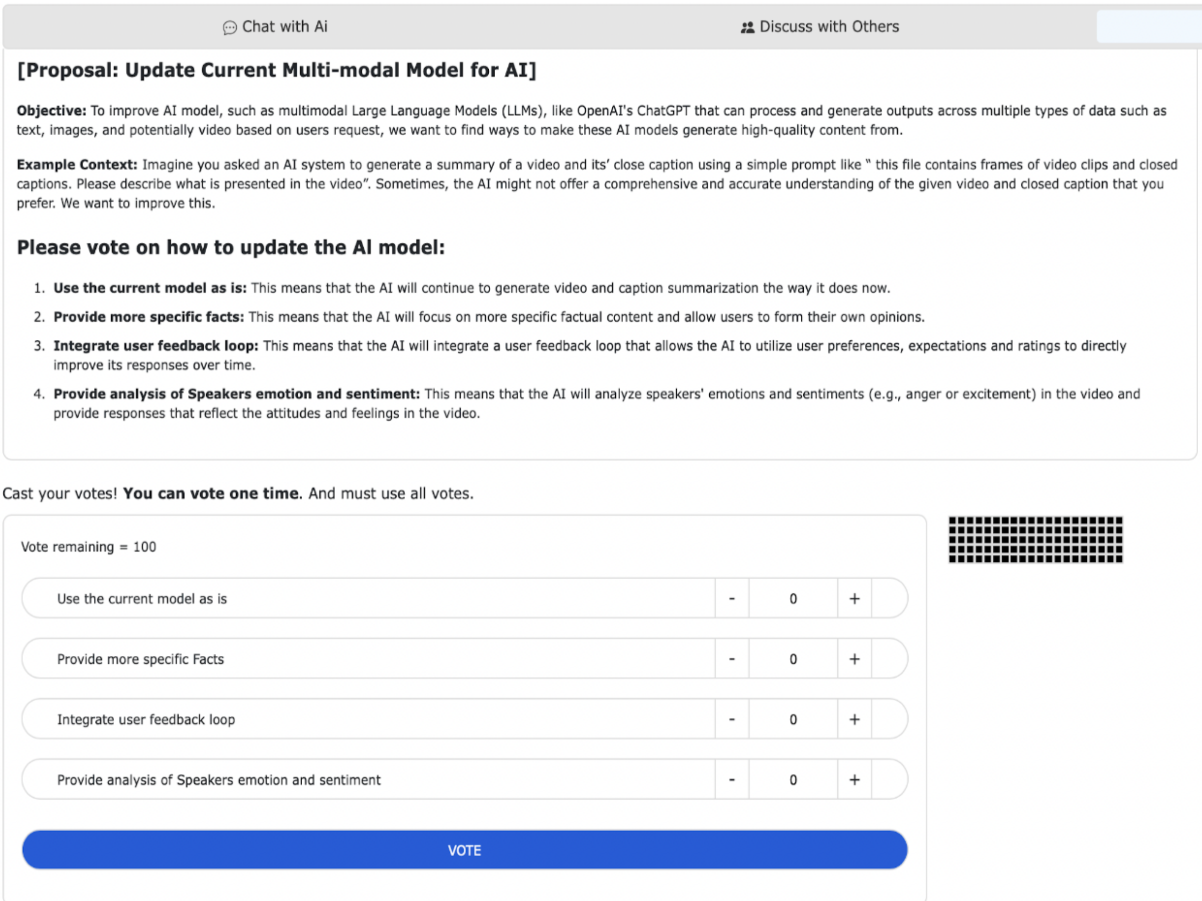


Figure 6: Governance decision making voting interface: Treatment condition example of quadratic voting interface for MM-llm decision for future improvement

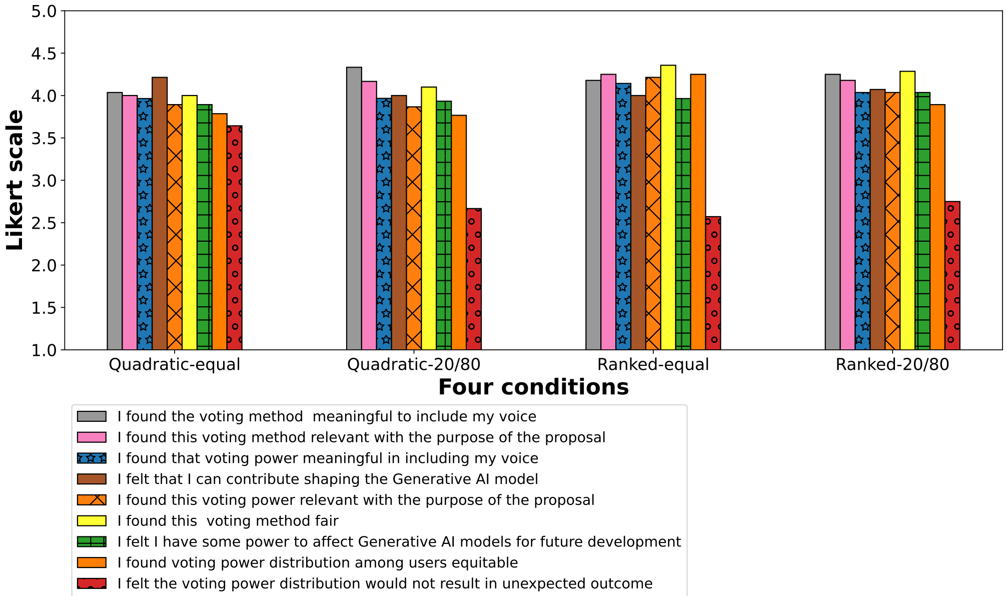


Figure 7: Users' perception of the quality of voting mechanism in governance decision making

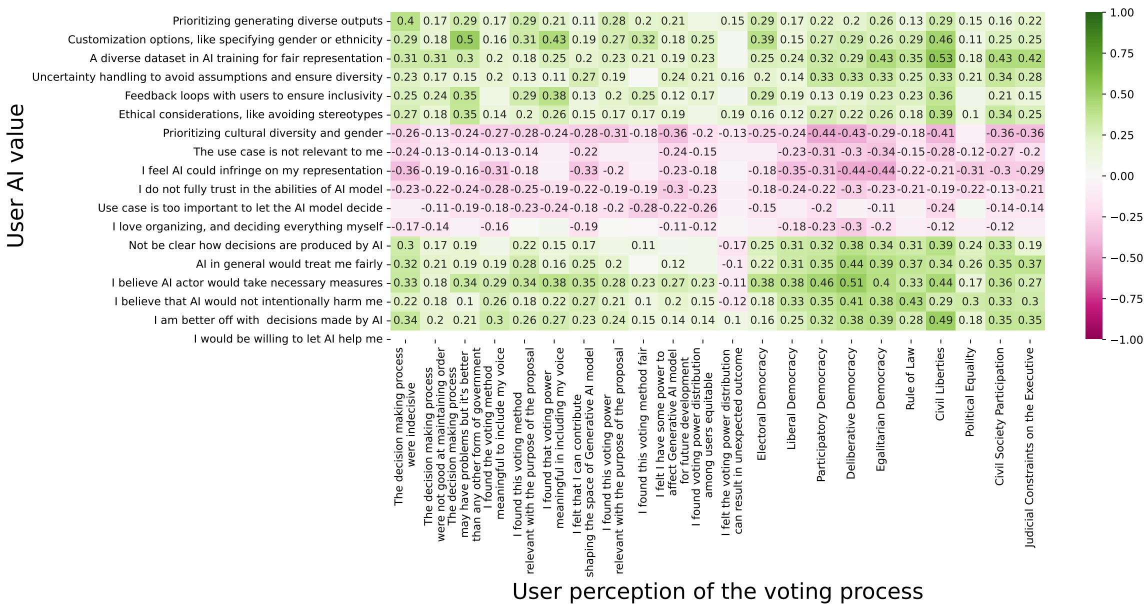


Figure 8: Correlation matrix of users' perceived quality of democracy (V-Dem Likert scale) with the predictor's variables that are users' perceived values on AI topics (Likert scale) including constructs, such as trust, perceived fairness, perceived accountability, and expected personalization.