

## RESEARCH ARTICLE

# AI and Deliberation: Normative Ideals in the Light of Current AI Research in the Context of Online Discussions—A Literature Review

Dennis Friess\*, Carina Weinmann† and Mira Warné†

This paper reviews current research on artificial intelligence (AI) in the context of online discussions, using deliberative norms as a theoretical framework. Focusing on deliberative dimensions of civility, equality, rationality, and reciprocity, the study examines and assigns 171 articles to these categories, finding that most AI research to date emphasizes enhancing rationality and civility in online discourse. Techniques such as argument mining and hate speech detection are commonly used to improve discussion quality, fostering more structured and respectful communication. However, efforts to address equality and reciprocity—the foundational principles of democratic discourse—remain sparse. This imbalance highlights critical gaps in the application of AI for fostering equitable and inclusive dialogue. The paper argues that current approaches often lack grounding in normative democratic standards, limiting their ability to promote genuine deliberation. By incorporating deliberative principles into AI design, researchers and practitioners can create tools that support democratic ideals, mitigating biases and amplifying underrepresented voices. Ultimately, this review provides a roadmap for leveraging AI to enhance civility, rationality, equality, and reciprocity in online discussions, paving the way for more inclusive and responsive digital public spheres.

**Keywords:** Literature Review; Artificial Intelligence; AI; Deliberation; Online Discussions

The field of artificial intelligence (AI) has seen remarkable growth in recent years, with a rapidly expanding body of literature exploring AI's applications in online discussions. Broadly defined, AI encompasses 'computational systems that involve algorithms, machine learning methods, natural language processing, and other techniques that operate on behalf of an individual to improve a communication outcome' (Hancock, Naaman & Levy 2020: 90). Thus, AI has introduced new methods for analyzing, guiding, and even shaping online interactions, making it a powerful tool in the context of digital discourse studies. Of particular interest are AI methods that intervene in online discussions, given the well-documented issues surrounding the low, and at times toxic, quality of these discussions (Coe, Kenski & Rains 2014; Wojcieszak et al. 2021).

Against this backdrop, the dual role of AI in online discussions is noteworthy: While the spread of AI-driven algorithms has been partly blamed for amplifying issues like echo chambers and polarizing content (Coleman 2018; Pariser 2011), AI is also increasingly viewed as a tool that might elevate online discussions and foster a more democratic debate (Rodríguez-Ruiz et al. 2020; Stoll,

Ziegele & Quiring 2020). This duality has spurred a range of research questions within communication studies, particularly concerning how AI can either hinder or enhance the quality of online discussions. Despite growing interest, much of the existing research on AI and online discussions remains focused on descriptive or technical aspects, largely addressing how AI functions within these settings or how it influences user behavior. What is often missing, however, is a critical examination of AI against normative standards rooted in democratic theories, specifically deliberative theory. Deliberative theory is one of the most influential perspectives in digital discourse studies (Carstens & Friess 2024), asserting that public discourse should meet certain normative standards—such as civility, equality, rationality, and reciprocity—to support democratic ideals (Friess & Eilders 2015). Yet, few studies have systematically analyzed AI's role in online discussions from this normative standpoint.

This paper seeks to bridge this gap by providing a comprehensive review of current AI research across different disciplines through the lens of deliberative norms. The guiding research question, therefore, is: *Which deliberative norms are addressed by research on the application of AI in online discussions to date?* Addressing this question requires not only examining how AI currently operates in online discussions but also investigating the extent to which AI supports or undermines the values essential to democratic discourse.

---

\* Institute for internet and democracy, University of Düsseldorf, DE† Institute for Social Science, Heinrich-Heine-University  
Düsseldorf, DECorresponding author: Dennis Friess ([dennis.friess@hhu.de](mailto:dennis.friess@hhu.de))

To answer this question, we will examine previous research against the background of the principles of civility, equality, rationality, and reciprocity, and assign each AI application to at least one of these principles. Accordingly, this review will identify where AI has been applied to uphold deliberative norms, and will also highlight areas where AI applications fall short. Thus, this analysis aims to provide meaningful starting points for both researchers and practitioners interested in advancing democratic discourse through AI. Ultimately, this article contributes to the dialogue on AI in communication studies by offering a normative perspective of current research, thus highlighting not only the achievements within the field but also the open questions and research desiderata that remain.

### Theoretical Framework

The transformation of public communication evoked by technological innovation has always preoccupied communication science (Dahlgren 1995; van Dijck & Poell 2015). However, political communication research has also been criticized for evaluating such processes of change retrospectively rather than actively shaping them (Schäfer & Wessler 2020). With regard to AI, we agree with Schäfer and Wessler (2020) that political communication scholars should critically analyze and evaluate the potential of AI for public communication—in this specific case, the potential for online publics—ex ante and based on theoretical and empirical expertise.

Some researchers have already looked theoretically at the question of how AI will impact democracy and the public sphere. Jungherr (2023) argues that in order to assess the influence of AI on democracy, the respective level must be considered, from the individual (micro level) to groups and institutions (meso level) to the entire public sphere (macro level). At the meso level, he sees equality being influenced above all by the fact that AI makes certain groups more or less visible due to its dependence on data. In another article, Jungherr and Schroeder (2023) argue that there are three main categories when it comes to the influence of AI on the public sphere: AI shapes information and behavior, it generates content, and it communicates itself. Based on these systematizations of AI's influence on democratic societies, we can classify our approach and aim of this article: Because it primarily affects different groups and institutions, the influence that AI exerts on online discussions may be subsumed under the meso level. With regard to the nature of the influence on online discussions, all three categories (i.e., shaping of information and behavior, generation of content, and communication) are probably relevant.

Nevertheless, although these systematizations offer useful heuristics for classifying the opportunities and risks that AI poses for democracy, they do not allow for a critical assessment. However, normative theories of democracy can do just that. They can not only assess various risks to democracy, such as those undoubtedly posed by generative AI in elections (Jungherr, Rauchfleisch & Wuttke 2024; Srivastava, Nikolich & Koppel 2023), from a sense of unease, but also substantiate them based on normative criteria. In addition, normative standards, such as those

established by deliberative theories, are, in our view particularly, well suited not only to assess the potential of AI for online public discourse, but also to actively shape its development, due to their paradigmatic status in the field (Friess & Eilders 2015).

The concept of deliberation originates from the literature on deliberative democracy (Gutmann & Thompson 2004; Habermas 1996). In essence, deliberative democrats argue that a certain type of communication, namely deliberation, can lead to several democratically desirable outcomes, such as increased political tolerance and participation (Mutz 2008). While there are many readings of the theoretical literature and competing conceptions of what exactly constitutes deliberation (Bächtiger et al. 2010), most authors share the idea that deliberation is a constructive and respectful exchange of reasons among equal participants (Friess & Eilders 2015). Following this minimal consensus, this study will define deliberation as a demanding type of communication characterized by the norms of civility, equality, rationality, and reciprocity.

(1) *Civility* refers to respectful interaction among discussion participants. This dimension is crucial because mutual respect creates a constructive and supportive discussion environment, even when disagreements arise (Mansbridge et al. 2012). When participants engage in a civil way, it minimizes hostility and encourages open dialogue, allowing everyone to voice their perspectives without fear of personal attacks or negative judgments. Civility, therefore, is fundamental to fostering an inclusive atmosphere conducive to democratic conversation (Papacharissi 2004). However, empirical studies have not only shown that online discussions tend to produce high volumes of hateful and derogatory user comments (Coe, Kenski & Rains 2014), but that this sort of communication can also negatively impact their readers in several ways, such as increased perceptions of bias or antisocial behavior (Anderson et al. 2018; Ziegele, Koehler & Weber 2018). Consequently, developing effective strategies to enhance the quality of user comments has become a vibrant research area.

(2) *Equality* means that all participants have equal rights and genuine opportunities to participate (Gutmann & Thompson 2004). From a deliberative perspective, this dimension is essential because it ensures that every voice can be heard and thus has the potential to shape the outcome (Habermas 1990). When equality is prioritized, it prevents dominant voices from overshadowing others, allowing a diversity of opinions to be considered (Bächtiger & Beauvais 2016). Yet, research shows that online deliberation often falls short in achieving equality, as the 'digital divide' or 'participatory divide' prevents certain groups from engaging in discussions online (Hargittai & Walejko 2008; Norris 2012). However, these divides primarily relate to access or structural equality (Graham 2008). By contrast and going one step further, discursive equality centers on the communication process itself. Graham (2008: 21) defines discursive equality as the principle that 'all participants within the process of deliberation be considered equal members', ideally resulting in an equitable distribution of voices. However, studies reveal that online discussions are frequently dominated by a small number of highly active

users (Graham & Wright 2014), which runs counter to the principle of equality.

(3) *Rationality* in deliberation implies that the discussion is grounded in reasoned arguments and the search for constructive solutions (Gutmann & Thompson 2004). This dimension is important because it directs the conversation towards an evidence-based exchange and thus unfolds the legitimizing power of deliberation (Manin 1987). When participants provide justifications for their views, contest other arguments, and propose practical solutions, deliberation unfolds its 'truth-tracking potential' (Habermas 2006: 413) and can lead to better-informed decision-making. Even though different authors have emphasized that there are alternative ways of reasoning, for example, through the expressions of emotions, narratives, and testimony (Dryzek 2002), the classic idea in which deliberation follows the ideal of rational discourse in which verifiable reasons are mutually contested is still very much dominant. Empirical research provides mixed evidence with regard to rationality in online discussions. While some studies find a fair amount of reasoning online (Berg, Fagerholm & Strandberg 2024), other findings provide a rather skeptical view, mostly regarding the complexity of argumentation provided in online discussions (Santana 2019). Another strand of literature suggests that rationality is highly contingent upon the design and context of online discussion environments (Esau, Friess & Eilders 2017; Marzinkowski & Engelmann 2022).

(4) *Reciprocity* is a foundational principle of deliberation and means that participants are supposed to refer to each other's statements, engaging in a back-and-forth exchange (Morrell 2018). Reciprocity is critical for deliberation's epistemic functions—facilitating learning, changing opinions, and fostering informed reasoning (Estlund & Landemore 2018). Engaging with diverse perspectives also enhances error prevention (Bohman 2007) and enables empathy, mutual understanding, and a shared focus on the common good (Barber 1984). In addition to that, reciprocity upholds norms of politeness and civility, promoting mutual recognition, community, and trust (Bächtiger & Beauvais 2016). Thus, Esau and Friess (2022) conclude that within deliberative theory, the role of reciprocity can hardly be overestimated. Against the backdrop of its theoretical importance, many studies have investigated reciprocity in online discussions (Aragón et al. 2017; Graham 2008). Studies suggest a general lack of reciprocity in online discussions, meaning that they more often take the form of monologic statements rather than reflexive discourses; many comments do not receive any sort of response (Esau & Friess 2022). Previous research indicates that the distribution of reciprocity is highly uneven: The frequency of social interactions often follows a law of power distribution, meaning that a small number of actors receive a large and disproportionate share of replies (Raban & Rabin 2009).

The concept of deliberation has been used widely to evaluate the democratic potential of online discourse. We think that the four normative dimensions discussed offer a suitable theoretical framework to review current AI research more systematically. Current research reveals

significant challenges in each area, indicating that while online discussions have the potential to foster democratic values, they frequently fall short in practice, often due to uncivil forms of discussion, unequal participation, limited rational engagement, and lack of reciprocity. As AI continues to shape the dynamics of online discourse, applying these deliberative principles will be crucial for highlighting how AI intervenes in democratic communication, identifying both the opportunities and potential risks.

## Method

The application of AI in the context of online discussions is a highly interdisciplinary and emergent field, encompassing contributions from computer science, communication studies, political science, and beyond. As such, it consists of a wide range of study types—ranging from conceptual discussions and normative proposals to empirical evaluations of AI applications. Our objective is to generate a broad overview of how AI has been proposed or implemented in relation to deliberative processes. Given the exploratory nature and conceptual breadth of our research interest, we adopted a *scoping review* as the appropriate methodological approach since it is particularly well-suited for mapping heterogeneous research areas and for identifying gaps, thematic patterns, and future research directions (Paré et al. 2015). Following the methodological guidelines for scoping studies by Arksey and O'Malley (2005), we structured the literature review along five stages.

The overarching aim of the review (*Stage 1*) was guided by a normative understanding of deliberation (see Theoretical Framework), which provided a conceptual lens for evaluating AI applications. Our guiding objective therefore is to explore in which areas—and to what extent—AI is being proposed or implemented in ways that intersect with deliberative ideals. We adopted a multi-pronged strategy to identify relevant studies (*Stage 2*). This included a keyword-based approach and Google Scholar as a starting point, followed by a snowballing process. The initial keyword combinations<sup>1</sup> covered terms related to 'artificial intelligence', 'deliberation', and 'online discussion' in both English and German (as the two languages the authors speak). To ensure topicality, we limited ourselves to articles published no earlier than 2015 whenever possible, although earlier work (identified by looking through reference lists) was included in cases of particular technical relevance or uniqueness of the AI task examined. Our inclusion criteria were intentionally broad to account for the interdisciplinary nature of the topic. To provide a comprehensive overview, we have not limited our search to specific disciplines, included gray literature such as policy reports and white papers, and included papers that do not explicitly refer to deliberation as a theoretical concept. As a result, the context was expanded, and papers were also included that deal with the use of AI in a broader sense of online discussions (including social media). However, for reasons of time and resources, only materials in English or German were considered. In total, 407 potentially relevant articles were identified based on the title (and abstract if the title was ambiguous) in June and July 2022.

All the articles were manually reviewed (*Stage 3*) to exclude literature that does not thematize the use of AI in the context of deliberation (in the broadest sense) and accordingly does not fit the research desideratum. This left us with 171 articles, which constitute the literature corpus. Despite the search not focusing on specific disciplines, these are primarily works from the field of computer science, but also from the social sciences and the field of computational social science. The articles were manually categorized by the authors using the theoretical framework presented above (*Stage 4*). After scanning the entire article, the coders decided whether the study—explicitly or implicitly—targets concepts related to one or more of the four key dimensions of deliberation, i.e., equality, rationality, civility, or reciprocity. In cases where AI applications did not clearly align with any of the four dimensions of deliberation but the article still addresses the use of AI to support online discussions, the articles were included in a separate category ('other'). Where applicable, we assigned additional tags to specify the particular focus within the dimensions. These tags were later used to inductively build subcategories. The complete list of included studies and their classifications is provided in Online Appendix 2. An article that targets more than one of the main dimensions or more than one specific aspect within a dimension may therefore appear more than once in the evaluation. In addition to that, a short note on the research design, the research question, the hypotheses, as well as the findings was added. The findings of the review are presented in the following section (*Stage 5*), organized thematically along the four deliberative dimensions. For each dimension, we provide illustrative examples of AI applications identified in the literature.

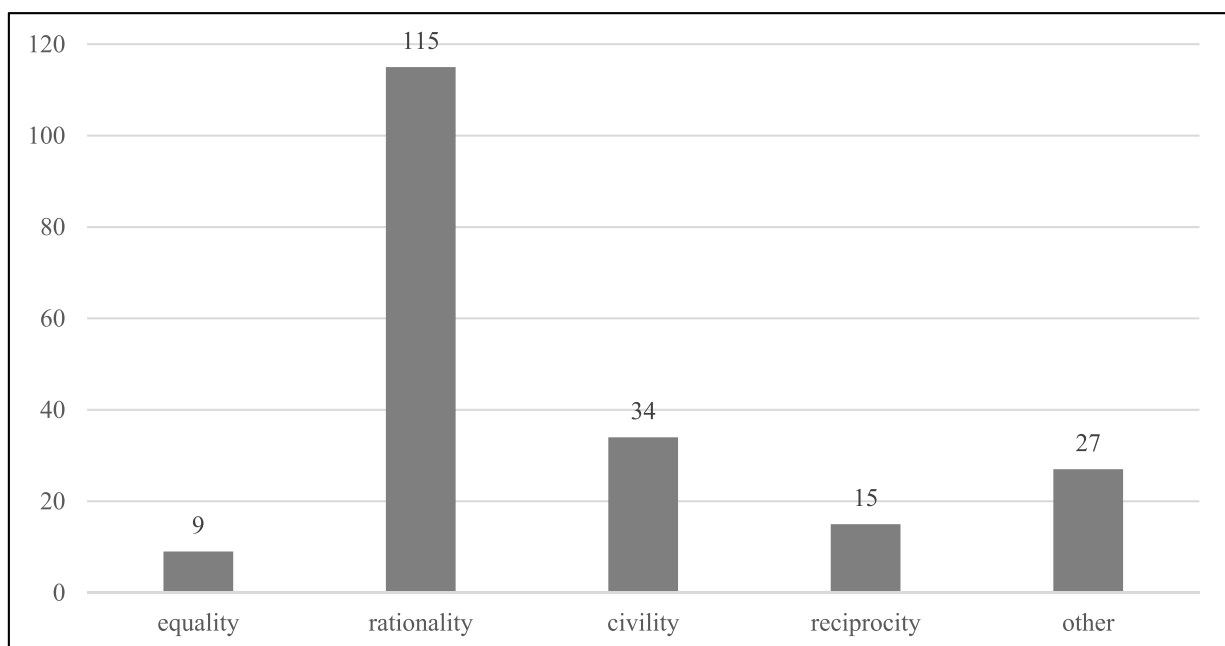
## Results

The results indicate that previous AI research has mainly focused on the dimensions of rationality and civility, whereas equality and reciprocity have received less

attention in AI research to date (**Figure 1**). The vast majority of the literature focuses on problems in online discussions that can be assigned to the dimension of *rationality* (115 articles). The main suggestion in this context is to use AI to improve the structural clarity of the discussion to make it more accessible to participants.

However, we have identified various solutions to this challenge, which justifies a closer look (**Figure 2**). Within the dimension of rationality, the largest proportion of studies (66 articles) addresses the complexity of online discussions by starting with increasing the transparency and comprehensibility of discussion structures and exploring the relationships between multiple arguments. A common starting point for this is argument-mining techniques. By automatically identifying discourse units and key components of arguments and then predicting agreement and contradiction, studies work on uncovering the relationships between arguments (Chakrabarty et al. 2020; Liebeck, Esau & Conrad 2017). This approach can also help make sense of large debates by identifying the issues that are central, controversial, or popular in a debate (Lawrence et al. 2017). Classifications through argument-mining techniques and argument clustering are further utilized to link or group related arguments (in terms of their stance or similarity) (Reimers et al. 2019). For example, Uszkoreit et al. (2018) implement a feature in their debate platform that allows users, when they want to make a new contribution, to automatically check whether an identical or similar argument already exists in the debate. Insights into argument structures and argument relationships can also be used to improve automated summarizations of debates (Fabbri et al. 2021).

Another proposed method for modeling discussion structures is the issue-based information system (IBIS). Compared to argument mining, this method includes an issue element and is further used as a basis for automated discussion facilitation (Gu et al. 2021; Ito et al. 2020). Other studies in this subfield focus on topic



**Figure 1:** Distribution of the articles across the four dimensions of deliberation.



analysis to detect emerging topics in online discussions (Mutanga & Abayomi 2022; Peng, Xu & Gan 2021). Topic modeling can further be used to promote structural clarity in discussions by grouping and structuring posts according to their topics (Liebeck, Esau & Conrad 2017), or by tracking changes in prevalent topics and structures over time (Stokes et al. 2020). Ultimately, the described techniques used for detecting underlying discussion structures can be used to visually promote structural clarity by implementing them on discussion platforms in the form of visual debate summarizations (Ullmann, de Liddo & Bachler 2019), argumentation maps, and discussion trees (Dragoni et al. 2018).

The second largest group (46 articles) comprises works that focus on the level of individual contributions. Since individual contributions are upstream of discussion structures, there are overlaps between the two areas, and therefore some articles, like some of the previously mentioned, cover both. Proposed AI techniques for analyses on the individual contribution level mainly group around the recognition of arguments and topics within a contribution. Specified tasks in this area encompass the recognition of a contribution stance (Sobhani, Inkpen & Matwin 2015), stance polarity (Sirrianni, Liu & Adams 2021), and emotional sentiment (Pandey 2021). Further proposals are dedicated to the automatic extraction of relevant keywords (Ito 2018), fact recognition and source identification (Dusmanu, Cabrio & Villata 2017), and detecting relevant answers to questions (Khan et al. 2020).

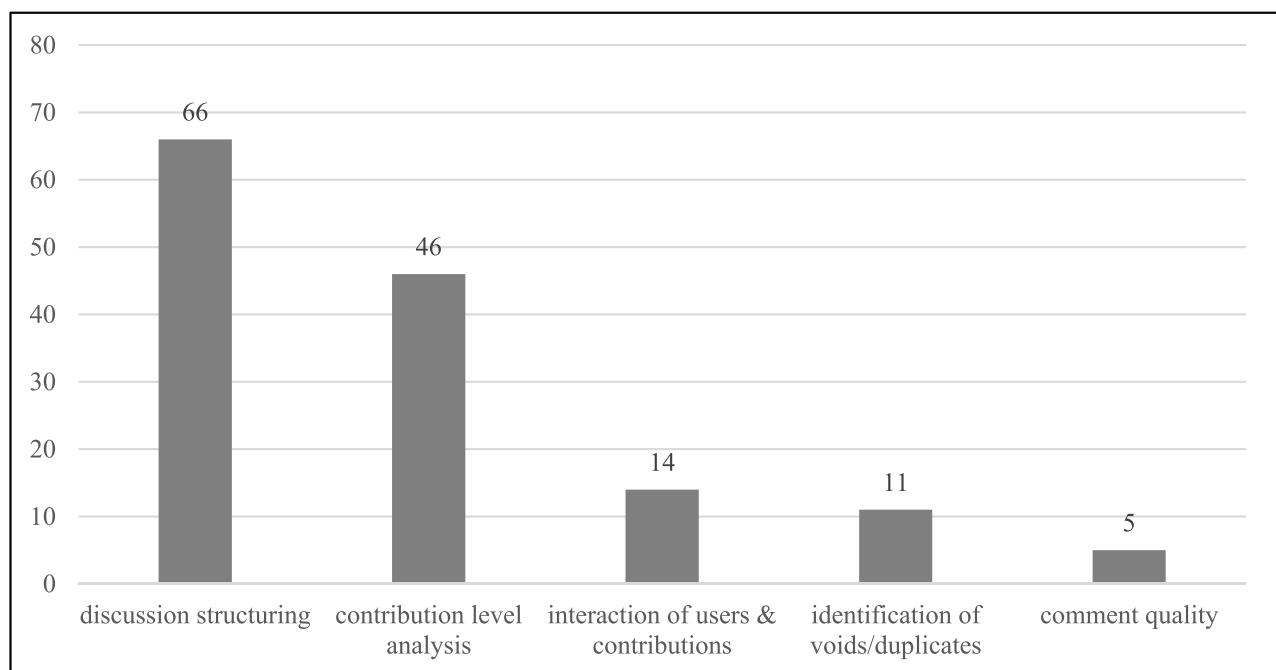
Far fewer studies (14 articles) attempt to deal with the connection between users and contributions (in a broader sense). The variety here ranges from connecting users with shared interests or opinions through user clustering or identifying subgroups (Arana-Catania et al. 2021; Qiu, Yang & Jiang 2013) to personalized summarizations of discussions based on individual interests (Tepper et al. 2018) to developing support for facilitation in the sense

of providing information about the dynamics between users and contributions. This also includes techniques for calculating the polarization of discourses, predicting collective opinions (Althuniyan et al. 2019), assessing individual impacts on debate outcomes (Mayfield & Black 2019), and evaluating both the persuasive power and the approval/rejection of particular arguments (Dutta, Das & Chakraborty 2020).

Equally few studies (11 articles) address the identification of surpluses (such as redundancies or duplicates), or conversely, the detection of unresolved aspects. For the automated detection of duplicates or redundancies, techniques such as clustering and semantic similarity detection using language models are used (Yang, Callan & Shulman 2006). To identify unresolved aspects, classification of discussions and their structure can help identify discussions with confusion or unanswered questions (Kim & Kang 2014).

Finally, studies proposing AI to check specific requirements for comments to promote the rationality of a discussion (five articles) are a marginal topic. Examples of AI interventions in this context include supporting moderation by automatically identifying posts that do not meet minimum formal requirements (e.g., regarding length or spelling) (Hudson, Archibald & Heap 2020) or those that do not comply with the thematic agenda (Vivod 2020).

Regarding the deliberative dimension of *civility* (34 articles), studies on the automated detection of contributions that violate this norm dominate the reviewed literature. The focus varies between different aspects of (in)civil behavior: Approaches for recognizing hate speech (Mozafari, Farahbakhsh & Crespi 2019), sarcasm (Verma, Shukla & Shukla 2021), or cyberbullying (Al-Garadi et al., 2019) are proposed. Some studies also combine these aspects and work toward a comprehensive detection of incivility (Stoll, Ziegele & Quiring 2020).



**Figure 2:** Distribution of the articles across the dimension of rationality.

Besides the pure detection of incivility, studies have also experimented with anticipating the occurrence of antisocial events (Zhang et al. 2018) or automated counter-speech bots (Clever, Klapproth & Frischlich 2022). Even though studies target different forms of deviant communication, this strand of the literature has advanced significantly in recent years and is already poised to contribute to more civil online discussions.

Compared to rationality and civility, the deliberative dimension of *reciprocity* (15 articles) is much less targeted by the AI research reviewed. Nevertheless, some promising attempts can be identified. Beyond the automated analysis of whether a contribution references other users' contributions (Morio & Fujita 2018), AI approaches are being pursued, which explore the potentials of bots or artificial moderators that encourage reference to other participants to increase reciprocity in online forums (Wyss & Beste 2017). Other studies have successfully confronted participants with contrary arguments that were algorithmically curated to encourage greater responsiveness (Krauthoff et al. 2018).

The deliberative dimension of *equality* (nine articles) is the least addressed by the reviewed research. We can therefore only mention individual works that aim to increase equal and balanced participation in the discourse, for example, through automated moderation bots that encourage participation (Kim et al. 2021). Another approach is the automated recognition of roles in the discussion, in particular the distinction between active and passive participants (Wijenayake et al. 2020).

Articles that we were not able to allocate within the framework (27 articles) and are therefore subsumed under 'other', for example, deal with the detection of social bots (Martini et al. 2021), misinformation (Hardalov et al. 2021) or work on the development of further support techniques for human moderators, such as real-time message recommendations (Lee et al. 2020).

## Discussion

This paper has dealt with the application of AI to online discussions, specifically with their use as a means to improve discussions in terms of the deliberative quality. To this end, we conducted a scoping review and categorized the AI applications suggested by existing work along four deliberative dimensions. Our findings reveal that the enhancement of rationality and civility in online discourse through AI can be identified as profound areas of research. Addressing the deliberative dimensions of equality and reciprocity through AI is not a blind spot, but research here is still in its infancy.

We can only speculate about the reasons for this imbalance. However, it seems likely that one reason is the complexity of the dimensions themselves. Equality and reciprocity are not only challenging from a theoretical point of view but also with regard to their practical operationalization that poses methodological challenges: Aspects such as equality and reciprocity often become apparent only through interactional patterns that unfold across a discussion rather than within single comments. Capturing them thus requires discourse-level annotation or dynamic contextual modeling, which are still difficult to

implement in current Natural Language Processing (NLP) and AI systems. This may explain why these dimensions have received comparatively little attention so far. At the same time, focusing primarily on normative ideals that are 'easily' measurable risks neglecting precisely those aspects—such as equality and reciprocity—that are complex but central to the deliberative ideal (Carstens & Friess 2024).

Furthermore, in the case of equality, there is also the fact that it is not possible to maintain or improve it in online discussions themselves. Rather, the degree of equality, especially with regard to the composition of the participants, is largely determined in the input dimension of online deliberation processes (Friess & Weinmann 2026), before the discussion even takes place, and is therefore beyond the reach of AI methods used in online discussions. Another reason is that we have looked at previous research through the deliberative lens, but very few studies do this themselves. We can even go a step further and state that many studies are not based on any theoretical background at all, but are instead motivated solely by practical considerations. This is probably due to the fact that although AI research has been going on for around 70 years, it has long been primarily motivated by engineering and computer science and has only recently found its way into communication science (Natale 2021). There may be papers from the field of communication science that theoretically discuss the use of AI in online discussions. However, the discipline has so far produced comparatively few concrete and functioning applications (but see Stoll, Ziegele & Quiring 2020), which is what we were looking for in our literature review. Accordingly, our literature corpus is dominated by papers from the field of computer science.

Nevertheless, we can conclude that our paper not only helps to provide a theoretically profound overview of the research to date but also shows where there are still gaps and further potential for using AI to improve online discussions from a normative democratic perspective. As the literature review has shown, one huge potential and, in our view, an urgent need to catch up exists with regard to the normative dimensions of equality. After all, the influence on equality, following Jungherr (2023), is one of the central ones with regard to democracy. This influence is due to the fact that existing biases, for example, the overrepresentation of certain minorities in crime statistics in data, can be perpetuated or even amplified by AI. However, this is precisely where AI could also come in, by specifically targeting such biases and attempting to overcome them. With regard to online discussions, for example, an AI application could identify minorities who tend to participate less in certain discourses, address them specifically, and encourage them to participate. By specifically addressing and encouraging discursive participation, the dimension of reciprocity could also be addressed and its extent increased in online discussions. In particular, an AI application that confronts participants in online discussions with opposing opinions could be useful by overcoming the well-known confirmation bias (Caraban et al. 2019). At the same time, it is worth considering whether the relative absence of such approaches in current AI research reflects a genuine gap

or rather the present limits of what is technically and normatively feasible. Addressing equality through AI in deliberative contexts may require a level of contextual understanding and ethical oversight that current systems cannot yet provide. Recognizing these constraints helps to distinguish between areas where progress is desirable in principle and those where caution remains warranted.

Furthermore, this paper provides an idea of how future research on AI can be democratically classified, justified, and evaluated, and thus joins current research that also pursues this goal (García-Marzá & Calvo 2024; Jungherr 2023). Finally, our study provides a starting point for ethical reflection by providing evidence that certain forms of communication are more likely to be affected by AI interventions than others. The potential fairness issues this raises have been recently discussed by Carstens and Friess (2024). They argue that the current focus on verbal markers of incivility and argument-mining approaches risks excluding minority voices and privileges those who have greater access to formal education and are thus well trained in rational argumentation. With this argument, they echo an already established critique of normative democratic theory (Young 1996). In addition, their article confirms the argument that AI not only replicates but could even reinforce biases with regard to minorities (Jungherr 2023).

In summary, this paper illustrates the opportunities and risks of AI for democracy and classifies them against the normative background of deliberative democracy. Using this background, it highlights significant progress in using AI to enhance rationality and civility in online discussions, while also revealing the need for more research on equality and reciprocity. By identifying these gaps, we provide a foundation for future studies to leverage AI in fostering more democratic online discourse.

## Conclusion

The rise of AI presents both promising and challenging dimensions for the future of democratic discourse online. Normative democratic theories, particularly deliberative democracy, provide crucial guidance in evaluating the extent to which AI can support democratic ideals and in identifying where AI may undermine democratic values, while allowing to advocate for solutions that uphold these principles. Going forward, interdisciplinary collaboration between AI developers, policymakers, and communication scholars will be essential in ensuring that AI serves as a force for inclusive, constructive, and democratic online public discourse.

## Note

- <sup>1</sup> A comprehensive list of the used Keywords is displayed in Online Appendix 1.

## Additional Files

The additional files for this article can be found as follows:

- **Online Appendix 1.** List of Keywords for the Scoping Review. DOI: <https://doi.org/10.16997/jdd.1805.s1>
- **Online Appendix 2.** List of Reviewed Publications and Tagged Dimensions. DOI: <https://doi.org/10.16997/jdd.1805.s2>

## Acknowledgments

We thank Maike Behrendt and Stefan Wagner for their valuable feedback and suggestions within the review process.

## Funding Information

This work was supported by the Jürgen Manchot Foundation in the context of the Manchot Research Group ‘Decision-Making with the Help of Artificial Intelligence’ at the Heinrich Heine University Düsseldorf (Germany).

## Competing Interests

The authors have no competing interests to declare.

## References

- Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., Mujtaba, G., Chiroma, H., Khattak, H. A., & Gani, A. (2019). Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges. *IEEE Access*, 7, 70701–70718. <https://doi.org/10.1109/ACCESS.2019.2918354>
- Althuniyan, N., Sirrianni, J. W., Rahman, M. M., & Liu, X. (2019). Design of mobile service of intelligent large-scale cyber argumentation for analysis and prediction of collective opinions. In D. Wang & L.-J. Zhang (Eds.), *Lecture Notes in Computer Science. Artificial intelligence and mobile services – AIMS 2019* (Vol. 11516, pp. 135–149). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-23367-9\\_10](https://doi.org/10.1007/978-3-030-23367-9_10)
- Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, 30(1), 156–168. <https://doi.org/10.1093/ijpor/edw022>
- Aragón, P., Gómez, V., & Kaltenbrunner, A. (2017). Detecting platform effects in online discussions. *Policy & Internet*, 9(4), 420–443. <https://doi.org/10.1002/poi3.158>
- Arana-Catania, M., van Lier, F.-A., 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. DLI: <https://doi.org/10.1145/3452118>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Bächtiger, A., & Beauvais, E. (2016). Taking the goals of deliberation seriously: A differentiated view on equality and equity in deliberative designs and processes. *Journal of Public Deliberation*, 12(2). <https://doi.org/10.16997/jdd.254>
- Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M. R., & Steiner, J. (2010). Disentangling diversity in deliberative democracy: Competing theories, their blind spots and complementarities. *Journal of Political Philosophy*, 18(1), 32–63. <https://doi.org/10.1111/j.1467-9760.2009.00342.x>



- Barber, B. R.** (1984). *Strong democracy*. Los Angeles, CA: University of California Press.
- Berg, J., Fagerholm, A., & Strandberg, K.** (2024). Quality user-generated content? A case study of the quality of online news comments on the site of Finnish public service broadcaster Yle. *Journalism Practice*, 1–22. <https://doi.org/10.1080/17512786.2024.2329759>
- Bohman, J.** (2007). Political communication and the epistemic value of diversity: Deliberation and legitimation in media societies. *Communication Theory*, 17(4), 348–355. <https://doi.org/10.1111/j.1468-2885.2007.00301.x>
- Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P.** (2019). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). New York, NY: ACM. <https://doi.org/10.1145/3290605.3300733>
- Carstens, J. A., & Friess, D.** (2024). AI within online discussions: Rational, civil, privileged?. *Minds & Machines*, 34(10), 1–25. <https://doi.org/10.1007/s11023-024-09658-0>
- Chakrabarty, T., Hidey, C., Muresan, S., Mckeown, K., & Hwang, A.** (2020, April 30). *AMPERSAND: Argument Mining for PERSuasive oNline discussions*. <http://arxiv.org/pdf/2004.14677v1>
- Clever, L., Klapproth, J., & Frischlich, L.** (2022). Automatisierte (Gegen-)Rede? Social Bots als digitales Sprachrohr ihrer Nutzer\*innen. In J. Ernst, M. Trompeta, & H.-J. Roth (Eds.), *Interkulturelle Studien. Gegenrede digital* (pp. 11–26). Wiesbaden: Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-36540-0\\_2](https://doi.org/10.1007/978-3-658-36540-0_2)
- Coe, K., Kenski, K., & Rains, S. A.** (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10.1111/jcom.12104>
- Coleman, M. C.** (2018). Bots, social capital, and the need for civility. *Journal of Media Ethics*, 33(3), 120–132. <https://doi.org/10.1080/23736992.2018.1476149>
- Dahlgren, P.** (1995). *Television and the public sphere: Citizenship, democracy and the media*. London: Sage.
- Dragoni, M., Da Costa Pereira, C., Tettamanzi, A. G. B., & Villata, S.** (2018). Combining argumentation and aspect-based opinion mining: The SMACK System. *AI Communications*, 31(1), 75–95. <https://doi.org/10.3233/AIC-180752>
- Dryzek, J. S.** (2002). *Deliberative democracy and beyond*. Oxford: Oxford University Press. <https://doi.org/10.1093/019925043X.001.0001>
- Dusmanu, M., Cabrio, E., & Villata, S.** (2017). Argument mining on twitter: arguments, facts and sources. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2317–2322). Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1245>
- Dutta, S., Das, D., & Chakraborty, T.** (2020). Changing views: Persuasion modeling and argument extraction from online discussions. *Information Processing & Management*, 57(2), 102085. <https://doi.org/10.1016/j.ipm.2019.102085>
- Esau, K., & Friess, D.** (2022). What creates listening online? Exploring reciprocity in online political discussions with relational content analysis. *Journal of Deliberative Democracy*, 18(1), 1–16. <https://doi.org/10.16997/jdd.1021>
- Esau, K., Friess, D., & Eilders, C.** (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, 9(3), 321–342. <https://doi.org/10.1002/poi3.154>
- Estlund, D., & Landemore, H.** (2018). The epistemic value of democratic deliberation. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, M. Warren, G. Smith, & M. Setälä (Eds.), *The Oxford handbook of deliberative democracy* (pp. 112–131). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198747369.013.26>
- Fabbri, A. R., Rahman, F., Rizvi, I., Wang, B., Li, H., Mehdad, Y., & Radev, D.** (2021, June 2). *ConvoSumm: Conversation summarization benchmark and improved abstractive summarization with argument mining*. <http://arxiv.org/pdf/2106.00829v1>
- Friess, D., & Eilders, C.** (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339. <https://doi.org/10.1002/poi3.95>
- Friess, D., & Weinmann, C.** (2026). Making online participation more intelligent: Considerations on how artificial intelligence can support public online participatory processes. In E. Paulis, R. Kies & A. Östling (Eds.), *Public deliberation in the digital age. Platforms, participation, and legitimacy* (pp. 143–157). London: Routledge.
- García-Marzá, D., & Calvo, P.** (2024). *Algorithmic democracy: A critical perspective based on deliberative democracy*. Cham: Springer. <https://doi.org/10.1007/978-3-031-53015-9>
- Gouvea, L. C., Garcia, A. C. B., & Vivacqua, A. S.** (2019). Behavior indicators for sensemaking of online discussions. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 1366–1371). IEEE. <https://doi.org/10.1109/SMC.2019.8914182>
- Graham, T.** (2008). Needles in a haystack. *Javnost – The Public*, 15(2), 17–36. <https://doi.org/10.1080/1318322.2008.11008968>
- Graham, T., & Wright, S.** (2014). Discursive equality and everyday talk online: The impact of “superparticipants”. *Journal of Computer-Mediated Communication*, 19(3), 625–642. <https://doi.org/10.1111/jcc4.12016>
- Gu, W., Moustafa, A., Ito, T., Zhang, M., & Yang, C.** (2021). A case-based reasoning approach for supporting facilitation in online discussions. *Group Decision and Negotiation*, 30(3), 719–742. <https://doi.org/10.1007/s10726-021-09731-4>
- Gutmann, A., & Thompson, D. F.** (2004). *Why deliberate democracy?* Princeton, NJ: Princeton University Press.
- Habermas, J.** (1990). *Moral consciousness and communicative action. Studies in contemporary German social thought*. Cambridge, MA: MIT Press.
- Habermas, J.** (1996). Three normative models of democracy. In S. Benhabib (Ed.), *Princeton paperbacks*.



- Democracy and difference: Contesting the boundaries of the political* (pp. 21–30). Princeton, NJ: Princeton University Press.
- Habermas, J.** (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4), 411–426. <https://doi.org/10.1111/j.1468-2885.2006.00280.x>
- Hancock, J. T., Naaman, M., & Levy, K.** (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
- Hardalov, M., Arora, A., Nakov, P., & Augenstein, I.** (2021). A survey on stance detection for mis- and disinformation identification. <https://doi.org/10.48550/arXiv.2103.00242>
- Hargittai, E., & Walejko, G.** (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, 11(2), 239–256. <https://doi.org/10.1080/13691180801946150>
- Hudson, C., Archibald, A., & Heap, T.** (2020). Integrating an AI-driven discussion platform: The impact of platform on engagement and quality. *EDEN Conference Proceedings*(1), 117–126. <https://doi.org/10.38069/edenconf-2020-ac0009>
- Ito, T.** (2018). Toward agent-based large-scale decision support system: The effect of facilitators. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 351–360. <https://doi.org/10.24251/HICSS.2018.047>
- Ito, T., Hadfi, R., Haqbeen, J., Suzuki, S., Sakai, A., Kawamura, N., & Yamaguchi, N.** (2020). Agent-based crowd discussion support system and its societal experiments. In Y. Demazeau, T. Holvoet, J. M. Corchado, & S. Costantini (Eds.), *Lecture Notes in Computer Science. Advances in practical applications of agents, multi-agent systems, and trustworthiness. The PAAMS Collection* (Vol. 12092, pp. 430–433). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-49778-1\\_41](https://doi.org/10.1007/978-3-030-49778-1_41)
- Jungherr, A.** (2023). Artificial intelligence and democracy: A conceptual framework. *Social Media + Society*, 9(3), Article 20563051231186353. <https://doi.org/10.1177/20563051231186353>
- Jungherr, A., Rauchfleisch, A., & Wuttke, A.** (2024). *Deceptive uses of Artificial Intelligence in elections strengthen support for AI ban*. <https://doi.org/10.48550/arXiv.2408.12613>
- Jungherr, A., & Schroeder, R.** (2023). Artificial intelligence and the public arena. *Communication Theory*, 33(2–3), 164–173. <https://doi.org/10.1093/ct/qtad006>
- Khan, A., Ibrahim, I., Uddin, M. I., Zubair, M., Ahmad, S., Al Firdausi, M. D., & Zaindin, M.** (2020). Machine learning approach for answer detection in discussion forums: An application of big data analytics. *Scientific Programming*, 2020, 1–10. <https://doi.org/10.1155/2020/4621196>
- Kim, J [Jihie], & Kang, J.-H.** (2014). Towards identifying unresolved discussions in student online forums. *Applied Intelligence*, 40(4), 601–612. <https://doi.org/10.1007/s10489-013-0481-1>
- Kim, S., Eun, J., Seering, J., & Lee, J.** (2021). Moderator chatbot for deliberative discussion. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–26. <https://doi.org/10.1145/3449161>
- Krauthoff, T., Meter, C., Baurmann, M., Betz, G., & Mauve, M.** (2018). D-BAS – A dialog-based online argumentation system. In J. Lawrence, K. Budzyńska, & S. Modgil (Eds.), *Frontiers in artificial intelligence and applications: Volume 305. Computational models of argument: Proceedings of COMMA 2018* (pp. 325–336). Amsterdam: IOS Press.
- Lawrence, J., Park, J., Budzynska, K., Cardie, C., Konat, B., & Reed, C.** (2017). Using argumentative structure to interpret debates in online deliberative democracy and eRulemaking. *ACM Transactions on Internet Technology*, 17(3), 1–22. <https://doi.org/10.1145/3032989>
- Lee, S.-C., Song, J., Ko, E.-Y., Park, S., Kim, J [Jihee], & Kim, J [Juho]** (2020). SolutionChat: Real-time moderator support for chat-based structured discussion. In R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguy, P. Bjorn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). New York: ACM. <https://doi.org/10.1145/3313831.3376609>
- Liebeck, M., Esau, K., & Conrad, S.** (2017). Text Mining für Online-Partizipationsverfahren: Die Notwendigkeit einer maschinell unterstützten Auswertung. *HMD Praxis Der Wirtschaftsinformatik*, 54(4), 544–562. <https://doi.org/10.1365/s40702-017-0321-6>
- Manin, B.** (1987). On legitimacy and political deliberation. *Political Theory*, 15(3), 338–368. <https://doi.org/10.1177/0090591787015003005>
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. F., & Warren, M. E.** (2012). A systemic approach to deliberative democracy. In J. Parkinson & J. Mansbridge (Eds.), *Deliberative systems: Deliberative democracy at the large scale* (pp. 1–26). Cambridge: Cambridge University Press.
- Martini, F., Samula, P., Keller, T. R., & Klinger, U.** (2021). Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data & Society*, 8(2), 205395172110335. <https://doi.org/10.1177/20539517211033566>
- Marzinkowski, H., & Engelmann, I.** (2022). Rational-critical user discussions: How argument strength and the conditions set by news organizations are linked to (reasoned) disagreement. *Digital Journalism*, 10(3), 433–451. <https://doi.org/10.1080/21670811.2021.1957968>
- Mayfield, E., & Black, A.** (2019). Stance classification, outcome prediction, and impact assessment. In S. Volkova, D. Jurgens, D. Hovy, D. Bamman, & O. Tsur (Eds.), *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science* (pp. 65–77). Stroudsburg, PA: Association for

- Computational Linguistics. <https://doi.org/10.18653/v1/W19-2108>
- Morio, G., & Fujita, K.** (2018). Annotating online civic discussion threads for argument mining. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 546–553). IEEE. <https://doi.org/10.1109/WI.2018.00-39>
- Morrell, M.** (2018). Listening and deliberation. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, M. Warren, G. Smith, & M. Setälä (Eds.), *The Oxford handbook of deliberative democracy* (pp. 236–250). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198747369.013.55>
- Mozafari, M., Farahbakhsh, R., & Crespi, N.** (2019, October 28). A BERT-based transfer learning approach for hate speech detection in online social media. <http://arxiv.org/pdf/1910.12574v1>
- Mutanga, M. B., & Abayomi, A.** (2022). Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development*, 14(1), 163–172. <https://doi.org/10.1080/20421338.2020.1817262>
- Mutz, D. C.** (2008). Is deliberative democracy a falsifiable theory? *Annual Review of Political Science*, 11, 521–538. <https://doi.org/10.1146/annurev.polisci.11.081306.070308>
- Natale, S.** (2021). Communicating through or communicating with: Approaching artificial intelligence from a communication and media studies perspective. *Communication Theory*, 31(4), 905–910. <https://doi.org/10.1093/ct/qtaa022>
- Norris, P.** (2012). *Digital divide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164887>
- Pandey, C.** (2021). redBERT: A topic discovery and deep sentiment classification model on COVID-19 online discussions using BERT NLP model. *International Journal of Open Source Software and Processes*, 12(3), 32–47. <https://doi.org/10.4018/IJOSSP.2021070103>
- Papacharissi, Z.** (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259–283. <https://doi.org/10.1177/1461444804041444>
- Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S.** (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, 52(2), 183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Pariser, E.** (2011). *The filter bubble: What the Internet is hiding from you*. New York, Penguin Press.
- Peng, X., Xu, Q., & Gan, W.** (2021). SBTM: A joint sentiment and behaviour topic model for online course discussion forums. *Journal of Information Science*, 47(4), 517–532. <https://doi.org/10.1177/0165551520917120>
- Qiu, M., Yang, L., & Jiang, J.** (2013). Mining user relations from online discussions using sentiment analysis and probabilistic matrix Factorization. In L. Vanderwende, H. Daumé III, & K. Kirchhoff (Eds.), *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 401–410). Stroudsburg, PA: Association for Computational Linguistics. <https://aclanthology.org/N13-1041>
- Raban, D. R., & Rabin, E.** (2009). Statistical inference from power law distributed web-based social interactions. *Internet Research*, 19(3), 266–278. <https://doi.org/10.1108/10662240910965342>
- Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I.** (2019, June 24). *Classification and Clustering of Arguments with Contextualized Word Embeddings*. <http://arxiv.org/pdf/1906.09821v1>
- Rodríguez-Ruiz, J., Mata-Sánchez, J. I., Monroy, R., Loyola-González, O., & López-Cuevas, A.** (2020). A one-class classification approach for bot detection on Twitter. *Computers & Security*, 91, 101715. <https://doi.org/10.1016/j.cose.2020.101715>
- Santana, A. D.** (2019). Toward quality discourse: Measuring the effect of user identity in commenting forums. *Newspaper Research Journal*, 40(4), 467–486. <https://doi.org/10.1177/0739532919873089>
- Schäfer, M. S., & Wessler, H.** (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz. *Publizistik*, 65, 307–331. <https://doi.org/10.1007/s11616-020-00592-6>
- Sirrianni, J. W., Liu, X., & Adams, D.** (2021). Predicting stance polarity and intensity in cyber argumentation with deep bidirectional transformers. *IEEE Transactions on Computational Social Systems*, 8(3), 655–667. <https://doi.org/10.1109/TCSS.2021.3056596>
- Sobhani, P., Inkpen, D., & Matwin, S.** (2015). From argumentation mining to stance classification. In C. Cardie (Ed.), *Proceedings of the 2nd workshop on argumentation mining* (pp. 67–77). Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W15-0509>
- Srivastava, B., Nikolich, A., & Koppel, T.** (2023). AI and elections: An introduction to the special issue. *AI Magazine*, 44(3), 216–217. <https://doi.org/10.1002/aaai.12110>
- Stokes, D. C., Andy, A., Guntuku, S. C., Ungar, L. H., & Merchant, R. M.** (2020). Public priorities and concerns regarding COVID-19 in an online discussion forum: Longitudinal topic modeling. *Journal of General Internal Medicine*, 35(7), 2244–2247. <https://doi.org/10.1007/s11606-020-05889-w>
- Stoll, A., Ziegele, M., & Quiring, O.** (2020). Detecting impoliteness and incivility in online discussions. *Computational Communication Research*, 2(1), 109–134. <https://doi.org/10.5117/CCR2020.1.005.KATH>
- Tepper, N., Hashavit, A., Barnea, M., Ronen, I., & Leiba, L.** (2018). Collabot: Personalized group chat summarization. In Y. Chang, C. Zhai, Y. Liu, & Y. Maarek (Eds.), *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 771–774). New York, NY: ACM. <https://doi.org/10.1145/3159652.3160588>
- Ullmann, T. D., de Liddo, A., & Bachler, M.** (2019). A visualisation dashboard for contested collective intelligence learning analytics to improve sensemaking

- of group discussion. *RIED. Revista Iberoamericana De Educación a Distancia*, 22(1), 41–80. <https://doi.org/10.5944/ried.22.1.22294>
- Uszkoreit, H., Gabryszak, A., Busemann, S., & Steffen, J.** (2018). Transparente und inklusive Online-Debatten durch Sprachtechnologie. In A. Burchardt & H. Uszkoreit (Eds.), *IT für soziale Inklusion* (pp. 1–10). Berlin: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110561371-001>
- van Dijck, J., & Poell, T.** (2015). Social media and the transformation of public space. *Social Media + Society*, 1(2), 205630511562248. <https://doi.org/10.1177/2056305115622482>
- Verma, P., Shukla, N., & Shukla, A. P.** (2021). Techniques of sarcasm detection: A review. In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 968–972). IEEE. <https://doi.org/10.1109/ICACITE51222.2021.9404585>
- Vivod, J.** (2020, November 23). *Using Machine Learning and Natural Language Processing Techniques to Analyze and Support Moderation of Student Book Discussions*. <http://arxiv.org/pdf/2011.11712v1>
- Wijenayake, P., Silva, D. de, Alahakoon, D., & Kirigeegamage, S.** (2020). Automated detection of social roles in online communities using deep learning. In *Proceedings of the 3rd International Conference on Software Engineering and Information Management* (pp. 63–68). New York, NY: ACM. <https://doi.org/10.1145/3378936.3378973>
- Wojcieszak, M. E., Thakur, A., Ferreira Gonçalves, J. F., Casas, A., Menchen-Trevino, E., & Boon, M.** (2021). Can AI enhance people's support for online moderation and their openness to dissimilar political views? *Journal of Computer-Mediated Communication*, 26(4), 223–243. <https://doi.org/10.1093/jcmc/zmab006>
- Wyss, D., & Beste, S.** (2017). Artificial facilitation: Promoting collective reasoning within asynchronous discussions. *Journal of Information Technology & Politics*, 14(3), 214–231. <https://doi.org/10.1080/19331681.2017.1338175>
- Yang, H., Callan, J., & Shulman, S.** (2006). Next steps in near-duplicate detection for eRulemaking. In J. A. B. Fortes & A. Macintosh (Eds.), *dg.o '06: Proceedings of the 2006 national conference on Digital government research* (pp. 239–248). New York, NY: ACM Press. <https://doi.org/10.1145/1146598.1146663>
- Young, I. M.** (1996). Communication and the other: Beyond deliberative democracy. In S. Benhabib (Ed.), *Princeton paperbacks. Democracy and difference: Contesting the boundaries of the political* (pp. 120–135). Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9780691234168-007>
- Zhang, J., Danescu-Niculescu-Mizil, C., Sauper, C., & Taylor, S. J.** (2018). Characterizing online public discussions through patterns of participant interactions. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–27. <https://doi.org/10.1145/3274467>
- Ziegele, M., Koehler, C., & Weber, M.** (2018). Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media*, 62(4), 636–653. <https://doi.org/10.1080/08838151.2018.1532430>

**How to cite this article:** Friess, D., Weinmann, C., & Warné, M. (2025). AI and Deliberation: Normative Ideals in the Light of Current AI Research in the Context of Online Discussions—A Literature Review. *Journal of Deliberative Democracy*, 21(1), pp. 1–11. DOI: <https://doi.org/10.16997/jdd.1805>

**Submitted:** 30 January 2025

**Accepted:** 21 November 2025

**Published:** XX Month 202X

**Copyright:** © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



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

## Typesetting queries

1. The following items have been included within the reference list, but are not cited within the text. For each un-cited reference, please advise where it should be cited in the text, or confirm that it can be removed from the reference list.

Ref: Gouvea, L. C., Garcia, A. C. B., & Vivacqua, A. S. (2019).