

A Framework for AI-Supported Mediation in Community-based Online Collaboration

SOOBIN CHO, Human Centered Design & Engineering, University of Washington, USA
 MARK ZACHRY, Human Centered Design & Engineering, University of Washington, USA
 DAVID W. MCDONALD, Human Centered Design & Engineering, University of Washington, USA

Reference Format:

Soobin Cho, Mark Zachry, and David W. McDonald. 2025. A Framework for AI-Supported Mediation in Community-based Online Collaboration. arXiv:2509.10015 [cs.HC] <https://arxiv.org/abs/2509.10015>

1 Introduction: From Moderation to Mediation in Online Collaboration

Online spaces involve diverse communities engaging in various forms of collaboration, which naturally give rise to discussions, some of which inevitably escalate into conflict or disputes. To address such situations, AI has primarily been used for moderation, such as Reddit’s AutoModerator, which automatically filters content, enforces rules, and detects spam [11]. As AI capabilities have advanced, researchers have explored more sophisticated systems to support moderation, such as a visual system that helps Discord moderators monitor multiple channels and detect early signs of toxicity [7], and AI models tailored to specific Discord communities that identify user intentions and moderate toxic comments in context [3].

Moderation systems are important because they help maintain order. But common moderation strategies of removing or suppressing content and users rarely address the underlying disagreements or the substantive content of disputes. Moderation systems tend to be reactive and only intervene after visible conflict has already emerged, sometimes after the damage is done. As a result, the underlying conflicts are not guided toward constructive or mutually beneficial outcomes.

Mediation, by contrast, offers a fundamentally different approach. Instead of suppressing disagreement, it fosters understanding, reduces emotional tension, and facilitates consensus through guided negotiation. Mediation not only enhances the quality of collaborative decisions but also strengthens relationships among group members. For this reason, we argue for shifting focus toward AI-supported mediation in community-based online collaboration. We acknowledge that mediation can be time and effort intensive, which motivates the question of whether AI can be leveraged in a mediation system to reduce the overall burden.

Mediating group conflict in online collaboration through AI is a complex problem that is deeply intertwined with group dynamics, processes, and outcomes. In this context, AI no longer acts as a peripheral technical tool, but takes on a more central role within the group. Importantly, such engagement can open up new and more creative forms of collective practice.

2 Background and Related Work

2.1 Definition of Mediation and Mediator Actions

Mediation, originally rooted in the legal system as a form of alternative dispute resolution, is a facilitated process in which a neutral third party helps disputants voluntarily reach a mutually acceptable agreement without imposing a decision [18, 4, 10, 14, 12]. Some common styles of mediation include facilitative, evaluative, and transformative mediation, each distinguished by the mediator’s role and approach [18, 26].

While the specifics may vary by context, mediation typically proceeds through six stages: the mediator begins by introducing the process and building trust; the parties then present their perspectives; key issues are identified and structured; possible solutions are generated; these options are evaluated and refined; and finally, the process concludes with agreement or closure [2, 19]. Throughout these stages, mediators perform a wide range of actions to guide the dialogue and support resolution. For example, Branting et al. systematically categorized 30 types of mediator speech acts, such as acknowledgement, refocusing, fact-checking, and restating positions [4]. Similarly, Janier et al. identified six types of mediator tactics, including redirection, clarification of misunderstandings, and addressing negative collateral implications [12].

As seen above, existing research on mediation has focused on what mediators do, identifying how the process unfolds and categorizing the concrete actions involved.

2.2 AI Systems for Online Mediation

In legal contexts, AI-based mediation systems have primarily appeared in the context of Online Dispute Resolution (ODR). ODR is a form of dispute resolution that leverages the strengths of the Internet, where technology acts as a “fourth party” that works alongside the mediator [22]. In ODR, AI performs specific tasks such as analyzing outcomes of similar past cases, drafting legal documents and presentations, generating joint agreements, and assisting in legal research [21, 1]. For example, SAMA in India analyzes case data, predicts outcomes, and suggests possible resolutions [15]; +Acordo, a Brazil-based platform, collects and analyzes claim details using pre-defined parameters to propose settlements [20]; and Rechtwijzer in the Netherlands gathers participants’ stance and priorities to identify points of agreement and recommend solutions [25].

In the HCI domain, researchers have explored how AI systems can support online group discussions as both a mediator and a facilitator. Some systems focus on conversational guidance, as demonstrated by Chen et al.’s study, in which AI steps in during ad hoc team discussions through actions such as asking for clarification, slowing down hasty decisions, restoring speaking turns, and introducing alternative perspectives [5]. Others place greater emphasis on participation balancing. Kim et al. and Do et al. developed group chat agents that manage time and actively prompt low-contributing members to participate [8, 9, 16, 17]. Finally, AI has also been used to support consensus building. Tessler et al.’s Habermas Machine collected individual opinions from ad hoc groups on political topics and iteratively generated group statements that the group as a whole can agree on [24].

Across both domains, these systems are primarily designed to assist or automate mediator actions, using participant stance and discussion content as the main, and often only, source of information. Overall, unlike the action side, the informational side of mediation has been largely overlooked.

3 Suggesting Framework

The role of the mediator has been described as “communicating selected information back and forth” [23, 10], emphasizing the importance of their access to and understanding of information. However, as discussed earlier, prior work has focused largely on mediator actions, with little attention to the informational side of mediation.

This gap becomes especially problematic in community contexts, where mediation relies on a much more complex and layered understanding of information. In such settings, mediators must go beyond the surface-level content of discussions to grasp deeper contextual knowledge, including group histories and social dynamics. Ultimately, in community-based mediation, it is the depth of what the mediator knows about the group that makes effective mediation possible.

3.1 Insights from Real-World Cases

Reflections on real-world cases of group mediation highlight how effective mediation often requires information beyond the discussion content.

For instance, in a PhD milestone exam I attended as a student, one presenter faced a barrage of difficult questions from a faculty member, escalating tension in the room. The professor originally assigned to oversee the session was absent, and a staff member eventually stepped in only to announce that time was up. What was truly needed, however, was a mediator who understood not just the academic content, but the exam’s purpose, the student’s level of experience, and the faculty-student dynamic. Without such contextual knowledge, the situation could not be steered constructively, turning what could have been a productive dialogue into a demoralizing experience.

In another case, as a UX researcher I temporarily filled in for a product manager where I led a cross-functional meeting. During the meeting a conflict arose between designers and developers. Although I understood both design and development workflows, I hesitated to intervene because I lacked knowledge of the power dynamics and past practices, such as whose perspectives typically carried more weight or how such disputes were usually resolved. The discussion ended without resolution leaving everyone dissatisfied, and leaving behind unresolved tensions.

These examples illustrate the challenge of effective mediation. Mediation is not simply a process that is executed, it has topical, contextual, and cultural aspects that good mediators either know or can quickly grasp.

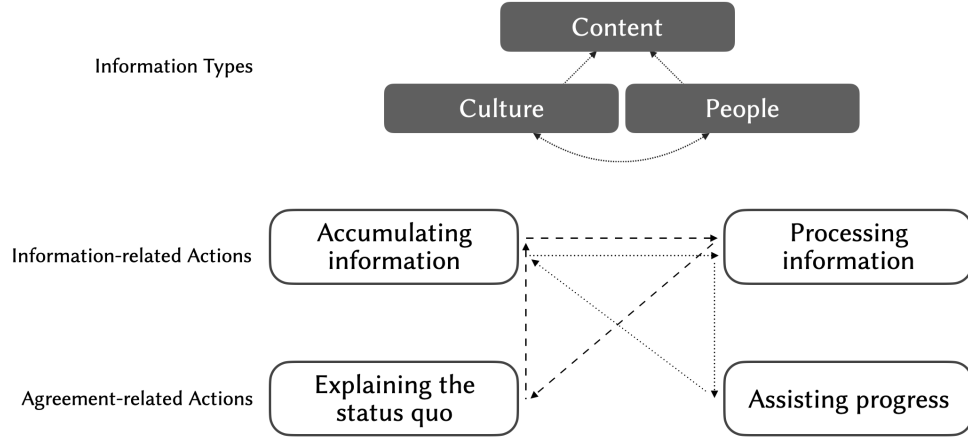


Figure 1: Information-focused framework for AI mediation in community-based collaboration.

3.2 Information-Focused Framework

Building on gaps left by prior research, we propose an information-focused framework for AI-supported mediation designed for community-based collaboration (Figure 1). Within this framework, we hypothesize that AI must acquire and reason over three key types of information: content, culture, and people.

Content refers to the topic under discussion and the claims being made. It is the most fundamental element of any discussion. To fully understand the content, however, it is also necessary to be familiar with both the culture and the people involved. *Culture* refers to the shared goals, norms, and expectations that shape how the discussion takes place. *People* refers to the participants’ experiences, roles, and relationships. Importantly, culture and people are mutually dependent: understanding the culture requires knowing who the participants are, and understanding the participants requires awareness of the cultural norms they operate within.

The mediator performs four actions across the three information types, including two information-related actions (accumulating and processing) and two agreement-related actions (explaining the status quo and assisting progress). The mediator operates through two recurring cycles of actions. The first cycle consists of accumulating and processing information to contextually explain the current status, while the second cycle consists of accumulating and processing information to assist progress. These two cycles alternate and together form a larger repetitive mediation process, where the outcome of each cycle becomes the basis for the next one.

4 Research Design and Trajectory

Our goal is to assess the feasibility of the proposed framework and demonstrate its potential as a foundation for designing AI mediation systems that support community-based online collaboration. To this end, we explore how the framework’s foundational elements operate in real-world online community settings.

We situate our study in Wikipedia, a well-established online community where collaboration continuously occurs around the creation of encyclopedic content. Wikipedia also has a long-standing and actively updated community culture that is extensively documented. It brings together contributors with diverse experiences and perspectives, whose personal histories and self-introductions are visible through public user pages. At the same time, prior research has noted that the absence of mediation in Wikipedia contributes to two major issues: a decline in the quality of collaborative decisions (i.e., article quality) and the deterioration of relationships among contributors (e.g., members leaving the community) [13].

Our research takes a phase-by-phase approach to examine how the proposed mediator actions should operate across the three types of information: content, culture, and people. So far, we have completed a study on how content information should be accumulated and processed, and we are currently investigating how it can be contextually explained.

The first study, accepted to *CSCW 2025*, involved 14 interviews with Wikipedians to understand how they make sense of disputes [6]. Each interview consisted of three parts: (1) reading an actual Wikipedia discussion using a

think-aloud protocol, (2) writing a summary of the discussion, and (3) reviewing a pre-generated AI summary of the same discussion. The findings revealed 13 key elements that participants considered important for understanding the discussion, such as usernames, evaluation of sources, editor behaviors, and rule violations. Participants noted that the AI summaries often failed to capture community-specific elements among these.

The ongoing study on content explanation involves a web-based prototype that displays a Wikipedia discussion thread with supporting tools. These tools provide various explanations of the discussion content, such as multiple types of summaries and extracted links. Through user interviews, we are examining how Wikipedians use these tools to understand discussions, what they find helpful, and what additional design features they would like to see.

5 Research Contribution

Our research clarifies and demonstrates the informational aspects of AI-supported mediation in community settings, which have been largely unaddressed in prior work. By taking into account complex contextual dynamics, our framework brings AI-supported mediation closer to the realities of communities and offers a more holistic perspective. In addition, the framework serves as a foundational guide regardless of the stage or scope of AI, whether it is implemented fully, partially, or in specific components.

References

- [1] Ryan Abbott and Brinson S. Elliott. 2023. Putting the Artificial Intelligence in Alternative Dispute Resolution: How AI Rules Will Become ADR Rules. *Amicus Curiae* 4 (2023), 685–706.
- [2] James Alfini, Joseph Stulberg, and Sharon Press. 2013. *Mediation Theory and Practice*. Vol. 3rd. Carolina Academic Press.
- [3] Henrik Axelsen, Johannes Rude Jensen, Sebastian Axelsen, Valdemar Licht, and Omri Ross. 2023. Can AI Moderate Online Communities? arXiv:2306.05122 [cs.CY] <https://arxiv.org/abs/2306.05122>
- [4] Karl Branting, Sarah McLeod, Sarah Howell, Brandy Weiss, Brett Profit, James Tanner, Ian Gross, and David Shin. 2022. A computational model of facilitation in online dispute resolution. *Artif. Intell. Law* 31, 3 (July 2022), 465–490. doi:10.1007/s10506-022-09318-7
- [5] XinHui Chen, Xiang Yuan, Hui Zhang, Ruixiao Zheng, and Wanyi Wei. 2025. Maintaining "Balanced" Conflict: Proactive Intervention Strategies of AI Voice Agents in Online Collaboration of Temporary Design Teams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 907, 19 pages. doi:10.1145/3706598.3713457
- [6] Soobin Cho, Mark Zachry, and David W. McDonald. 2025. Towards Insider Summarization for Mediation Instead of Moderation: Examining Wikipedian Views on Key Elements of Discussion Summaries. *Proc. ACM Hum.-Comput. Interact.* 9, CSCW500 (Nov. 2025). doi:10.1145/3757681
- [7] Frederick Choi, Tanvi Bajpai, Sowmya Pratipati, and Eshwar Chandrasekharan. 2023. ConvEx: A Visual Conversation Exploration System for Discord Moderators. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 262 (Oct. 2023), 30 pages. doi:10.1145/3610053
- [8] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Karrie Karahalios, and Brian P. Bailey. 2023. Inform, Explain, or Control: Techniques to Adjust End-User Performance Expectations for a Conversational Agent Facilitating Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 343 (Oct. 2023), 26 pages. doi:10.1145/3610192
- [9] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Jaewook Lee, and Brian P. Bailey. 2023. To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 99 (April 2023), 23 pages. doi:10.1145/3579532
- [10] Maria Goltsman, Johannes Horner, Gregory Pavlov, and Francesco Squintani. 2009. Mediation, Arbitration and Negotiation. *Journal of Economic Theory* 144 (07 2009), 1397–1420. doi:10.1016/j.jet.2008.08.010

- [11] Qinglai He, Kevin Hong, and Raghu Santanam. 2021. The Effects of Machine-powered Platform Governance: An Empirical Study of Content Moderation. *SSRN Electronic Journal* (01 2021). doi:10.2139/ssrn.3767680
- [12] Mathilde Janier and Chris Reed. 2017. Towards a Theory of Close Analysis for Dispute Mediation Discourse. *Argumentation* 31 (03 2017). doi:10.1007/s10503-015-9386-y
- [13] Dariusz Jemielniak. 2014. *Common Knowledge? An Ethnography of Wikipedia*. Stanford University Press, Stanford, California.
- [14] DEINIOL JONES. 2000. Mediation, conflict resolution and critical theory. *Review of International Studies* 26 (10 2000), 647 – 662. doi:10.1017/S0260210500006471
- [15] Karthikeyan C Karthikeyan C. 2024. *AI (Artificial Intelligence) for Conflict Resolution and Negotiation: Enhancing Mediation and Collaboration Through Intelligent Technology*. 21–50. doi:10.4018/979-8-3693-8442-8.ch002
- [16] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376785
- [17] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 87 (April 2021), 26 pages. doi:10.1145/3449161
- [18] Mykola Lohvinenko, Mykola Starynskyi, Lyudmila Rudenko, and Iryna Kordunian. 2021. *CONFLICT RESOLUTION QUARTERLY* 39, 1 (2021), 51–65. <https://search.informit.org/doi/10.3316/agispt.20210914053411>
- [19] Christopher W. Moore. 2014. *The Mediation Process: Practical Strategies for Resolving Conflict*. Vol. 4th. Jossey-Bass.
- [20] Isabela Motta, Julia Lima, Thomaz Ambrosio, Alimed Celecia, Ana Mangeth, Isabella Frajhof, and Cesar Cury. 2023. Towards designing Artificial Intelligence (AI)-based Online Dispute Resolution (ODR) services: The case of +Acordo. *Service Design and Innovation Conference*, 1193–1215. doi:10.3384/ecp203061
- [21] Orna Rabinovich-Einy and Ethan Katsh. 2021. *Online Dispute Resolution: Theory and Practice*. Eleven International Publishing, Chapter Artificial Intelligence and the Future of Dispute Resolution: The Age of AI-DR, Page Range.
- [22] Janet Rifkin. 2010. Online Dispute Resolution: Theory and Practice of the Fourth Party. *Conflict Resolution Quarterly* 19 (12 2010), 117 – 124. doi:10.1002/crq.3890190109
- [23] Linda Singer. 1990. *Settling Disputes: Conflict Resolution In Business, Families, And The Legal System*. Westview Press.
- [24] Michael Henry Tessler, Michiel A. Bakker, Daniel Jarrett, Hannah Sheahan, Martin J. Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C. Parkes, Matthew Botvinick, and Christopher Summerfield. 2024. AI can help humans find common ground in democratic deliberation. *Science* 386, 6719 (2024), eadq2852. arXiv:<https://www.science.org/doi/pdf/10.1126/science.adq2852> doi:10.1126/science.adq2852
- [25] John Zeleznikow. 2021. Using Artificial Intelligence to provide Intelligent Dispute Resolution Support. *Group Decision and Negotiation* 30 (04 2021), 789–812. doi:10.1007/s10726-021-09734-1
- [26] Zena Zumeta. 2000. Styles of mediation: Facilitative, evaluative, and transformative mediation. *National Association for Community Mediation Newsletter* 5 (2000).