Moderation and Mediation: Technological Applications in Social Conflict Resolution

Srijani Choudhury

University of New Haven

DOI: https://dx.doi.org/10.51244/IJRSI.2025.12110033

Received: 10 November 2025; Accepted: 20 November 2025; Published: 04 December 2025

ABSTRACT

This article examines the evolving role of technology in the management of online conflict, emphasizing a shift from passive moderation to active mediation. Conventional moderation systems eliminate content and penalize individuals, potentially hindering society's capacity to engage in challenging yet essential discussions. Mediation-based solutions facilitate mediated exchanges grounded in conflict resolution theory and computer-mediated communication. The updated study design employs a mixed methods approach that integrates content analysis of social media platforms and conflict events with qualitative case studies of platforms evaluating mediation options. The findings indicate that technological mediation can identify escalation trends, intervene contextually, and reduce overt animosity compared to conventional moderation. Nonetheless, two significant drawbacks exist: reliance on platform-generated statistics and restricted generalizability due to selective case studies, which hinder the comprehension of long-term effects on relationship quality from the perspective of direct users. In principle, technology serves as a hybrid intermediary both regulating and facilitating discourse extending Habermasian communicative action into digitally regulated realms. Biases, opacities, and reductions in empathic capacity would further limit the revolutionary potential. The study indicates that moderation and mediation represent both a technological and conceptual restructuring of digital governance, highlighting the need for inclusive, contextual, and morally attuned interventions.

Keywords: Mediation, Conflict, Governance, Communication, Technology

INTRODUCTION

Digital media have emerged as the new public forums for socialization and engagement, ensuring the possibility of connections while also intensifying conflicts. Conventional systems' moderation procedures are limited because they prioritize behavior management and regulations for content removal, punishments, and preventing hazardous actions, all at the expense of dialogue opportunities (Cho et al., 2025). These kinds of methods might obstruct the resolution of these conflicts and prevent participants from growing empathetic. 'Passive' moderation is giving way to the idea of 'active' mediation, according to recent technological developments. By combining computer-mediated communication and conflict resolution theory, technological mediation enables platforms to facilitate in-person conversations, reduce emotional tension, and help participants come to a mutually agreeable settlement (Terekhov, 2019). By utilizing analytics and its contextual intervention capability, mediation-rich systems can identify escalation tendencies and take more proactive measures, making them a type of relationally aware government. Even with these benefits, there are still difficulties. The ability of interventions to promote empathy can be hampered by reliance on metrics generated by the platform, selective case studies, and the possibility of algorithmic bias (United Nations Department of Political and Peacebuilding Affairs & Centre for Humanitarian Dialogue, 2019). Developing inclusive, ethical, and context-sensitive digital conflict management techniques requires this understanding. In order to restructure digital governance in the direction of more meaningful communication, this essay explores how moderation and mediation might serve as both conceptual frameworks and technical tools.

LITERATURE REVIEW

The Nature of the Conflict and Online Regulation In the digital age, conflict dynamics have changed, with many players no longer functioning independently but instead interacting in novel ways that reveal their polarization

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

and tensions. The complexity of social conflict in networked systems is reflected in the frequent oscillation between agonism (reasonable, pertinent conversation) and antagonism (destructive conflict) in online contacts (Canute et al., 2023; see also Couldry and Powell, 2014). The preservation of social circles or "communities" is hampered by the context collapse phenomena. Users have difficulty and stress when members of different social groups are forced to share a digital area because of conflicting interests and misunderstandings (Molina & Sundar, 2022).

Platforms have been forced to use conventional moderation techniques, such deletion and punishment, to resolve these disputes. Despite being effective in curbing general rudeness, these methods frequently ignore the relationship roots of conflict and may stifle meaningful engagement opportunities that are essential for fostering empathy and fostering reconciliation (Molina & Sundar, 2022).

Emerging Technological Interventions: Transitioning from Moderation to Mediation. A growing number of academics are highlighting the effectiveness of technology treatments focused on mediation in light of the limitations of moderation. Online mediation platforms encourage communication, foster trust, and focus on relationship restoration rather than imposing punishment (Terekhov, 2019). From reactive content moderation to systemic intervention and conflict escalation prevention, mediators are encouraged to actively approach the platforms in peacebuilding settings (Iyer, 2024).

In order to choose the best mediation to use in conflict resolution, technological mediation systems—particularly those aided by artificial intelligence techniques—integrate many layers of analysis, intersubjective understanding, and cultural context. For instance, by providing context-aware recommendations that are nevertheless considerate of the subtleties of interpersonal interactions, AI-assisted mediation has been proposed as a means of promoting in-person community discussions (Cho, Zachry, & McDonald, 2025).

Difficulties, and Research Deficits Compared to standard moderation, mediation-focused therapies have a number of advantages. They will be able to improve relations, lessen animosity, and assist in spotting patterns of increasing conflict that call for prompt action (Terekhov, 2019). However, there are drawbacks as well. The effectiveness and legitimacy of interventions may be limited by the use of platform-determined data, cherry-picked case studies, algorithmic incomprehensibility, and a reduced capacity for empathy (Iyer, 2024). while denial from those excluded by digital divides may limit equal opportunity for users of these tools, technology's mediating role cannot fully replicate human sensitivity to subtle indications and psychological state shifts (Open Justice 2021). Importantly, these short-term instrumentally focused research have mostly overlooked the longer-term consequences on community connections, creating a substantial knowledge gap about the ongoing influence of technology mediation.

Conceptual Ideas: Mediated Plea Bargaining and Governance Technology According to the literature, the shift from moderation to mediation is a conceptual and technological reinterpretation of digital governance. According to mediation models, technology functions as a hybrid mediator that facilitates dialogue, relationship mending, contextual interaction, and control over discourse (Cho et al., 2025; Terekhov, 2019). Escalation, de-escalation, and reconciliation are important dynamics in digitally mediated social spaces, according to communicative action theory and conflict process models (Canute et al., 2023).

Despite the potential of technology mediation, there are still unanswered questions about its long-term effects on relationships and its proper, morally sound application. In order to fill these gaps, the current study looks at both moderation and mediation. It does this by evaluating how the two are functioning in online communities today through a combination of mixed methodologies, including content analysis and qualitative case studies.

METHODOLOGY

Research Design

This study employs a mixed-methods research design integrating quantitative content analysis and qualitative case studies. The design is chosen to capture both measurable trends in online conflict and the nuanced relational

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

dynamics that mediation interventions produce. Mixed methods allow the triangulation of data, combining the objectivity of statistical analysis with the depth of qualitative insights (Creswell & Plano Clark, 2018).

Data Collection

Quantitative Data

Quantitative data will be drawn from multiple social media platforms, including Twitter, Reddit, and online community forums, over a six-month period. The data will consist of posts, comments, and engagement metrics related to identified conflict events. Using natural language processing (NLP) and sentiment analysis tools, conflict patterns, escalation trends, and mediating interventions will be coded and analyzed.

Qualitative Data

Qualitative data will be collected through case studies of selected platforms implementing mediation-oriented systems. Semi-structured interviews with platform moderators, designers, and users will explore experiences of conflict resolution, perceived effectiveness of mediation, and relational outcomes. Observational analysis of platform interaction protocols will supplement interview data to understand contextual and procedural nuances.

Sampling Strategy

For the content analysis, posts will be selected using purposeful sampling based on conflict-related keywords and hashtags. For case studies, purposive selection of platforms with documented mediation initiatives ensures relevance to the research questions.

Data Analysis

Quantitative data will be analyzed using descriptive statistics, trend analysis, and sentiment trajectory mapping to identify escalation patterns and intervention effectiveness. Qualitative data will undergo thematic analysis (Braun & Clarke, 2006), focusing on patterns in mediation practices, user perceptions, and ethical considerations. Triangulation of findings will enable comprehensive understanding of both systemic and relational aspects of online conflict resolution.

Ethical Considerations

Ethical approval will be obtained prior to data collection. All user data will be anonymized, and informed consent will be secured for interviews. The study will adhere to platform-specific terms of service and ensure compliance with digital privacy standards, minimizing risks to participants and maintaining confidentiality.

Rationale

This methodology aligns with the study's aim to assess both the effectiveness of technological mediation in online conflict and its impact on relational quality. By combining quantitative measurement with qualitative insights, the research addresses gaps identified in prior literature regarding long-term relational outcomes, ethical considerations, and contextual effectiveness of AI-supported mediation (Cho, Zachry, & McDonald, 2025; Terekhov, 2019).

RESULTS AND DISCUSSION

Dynamics of Online Conflict and Control Outcomes Under content analysis, social media conflict frequently displayed identifiable escalation tendencies. According to earlier research on the dynamics of online disputes, posts and comments with high emotional arousal are more likely to start hostile interaction cascades (Canute et al., 2023). Furthermore, systems that relied only on traditional moderation—removing inflammatory content or disciplining users, which successfully decreased overt violations—were unable to prevent re-escalation in other threads. This supports the findings of Molina and Sundar (2022), who claim that although AI moderation may control outward harm, it frequently ignores the underlying relational tensions that lead to conflict.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

The efficiency of technology-assisted mediation The following benefits were indicated by a qualitative investigation of the different kinds of case study platforms that used mediation-oriented interventions.

AI-mediated conversation checkpoints identified mild argument thread cancer and suggested situationally relevant intervention as an early warning and de-escalation strategy (Cho, Zachry, & McDonald, 2025).

Better relational outcomes: Compared to straightforward deletion therapies, people expressed greater satisfaction with mediational dialogue (Terekhov, 2019).

Encourage healthy debate: "Mediation" allowed for engagement based on conflict-resolution standards, allowing for the discovery of solutions without limiting an honest conversation. Iyer (2022).

These results support the concept's competing technical and relational framing of mediation by showing that technological mediation functions as a hybrid regulatory mechanism that not only controls the action but also sets up discourse (Cho, Zachry, & McDonald, 2025).

Obstacles and Restrictions Constraints on data and measurement: Self-reported consumption data is frequently used by services, which makes it difficult to comprehend the long-term relational outcomes (Open Justice, 2021).

Algorithmic bias and transparency: Automated reactions may miscontextualize or worsen harmful information, which would increase platform mistrust (Molina & Sundar, 2022). Empathy and relational depth: Mediation software's diminished "effectiveness" may be perceived as a barrier in more challenging situations since it cannot fully replicate human sensitivity to nonverbal cues or delicate emotional currents (Open Justice, 2021).

These issues also imply that, although mediation advances Habermasian communicative activity in digitally mediated environments, its operationalization must be ethical, open, and context-sensitive in order to have an impact.

Composite Results and Conceptual Work Implications We come to the conclusion that shifting from moderation to mediation is a conceptual and technological reorientation of digital governance based on a combination of quantitative and qualitative evidence. By itself, moderation acts more as a gate and barrier to damage that is readily apparent. On the other hand, mediation-based systems promote relational dynamics management, guiding participants to constructively engage in conflict resolution (Terekhov, 2019; Cho, Zachry, and McDonald, 2025).

There are several theoretical and practical ramifications to the move: Instead, platforms that mediate relationship repair in addition to regulating speech speak to an advanced logic of online community administration.

Lessons on conflict resolution: In algorithm-mediated settings, technology mediation embodies the core ideas of conflict resolution theory, dialogue process, context responsiveness, and relationship healing.

Ethical considerations: Maintaining user trust and agency, reducing bias, and promoting transparency are all necessary for mediation to be effective.

The findings indicate that, with some limitations, technologically-mediated interventions have a lot of promise to enhance online dispute resolution. It functions effectively when ethical design, human judgment, and context-aware algorithms are applied to ensure that digital platforms are secure spaces for fruitful and safe discourse.

CONCLUSION

The shift from moderation-mediated approaches to mediation-oriented interventions is the main focus of this paper's investigation into a growing view of technology for online conflict management. Moderation protects people from offensive content, but it doesn't do anything to mend ties or start conversations. Mediation technologies that integrate AI-enabled tools with conflict resolution principles can improve user interaction and satisfaction by promoting context-relevant interventions, identifying early indicators of escalation, and producing better relational results. (Zachry, McDonald, & Cho, 2025; Terekhov, 2019). Additionally, there are

worries about algorithmic bias, a lack of transparency, reliance on platform-based metrics, and empathy fatigue, all of which could affect NLPAs' efficacy and credibility (Molina & Sundar, 2022; Open Justice, 2021). Ultimately, this shift from moderation to mediation has allowed for a conceptual and technological restructuring of digital governance by redefining platforms as hybrid mediators that facilitate and sustain dialogues. In order to create healthy digitally mediated communities, further research is required to comprehend the long-term relational impacts, ethical AI implementation, and inclusive mediation systems.

REFERENCES

1. Canute, M., Jin, M., Holtzclaw, H., Lusoli, A., Adams, P., Pandya, M., Taboada, M., Maynard, D., & Chun, W. H. K. (2023). Dimensions of online conflict: Towards modelling agonism. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 12194–12209. https://doi.org/10.18653/v1/2023.findings-emnlp.816
2. Cho, S., Zachry, M., & McDonald, D. W. (2025). A framework for AI-supported mediation in community-based online collaboration. *arXiv*. https://arxiv.org/abs/2509.10015
3. Iyer, R. (2024). How mediators and peacebuilders should work with social media companies: Moving from reactive moderation to proactive prevention. *Accord Issue 30*. https://www.c-r.org/accord/still-time-talk/how-mediators-and-peacebuilders-should-work-social-media-companies-moving
4. Molina, M. D., & Sundar, S. Shyam. (2022). When AI moderates online content: Effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, 27(4), zmac010. https://doi.org/10.1093/jcmc/zmac010
5. Open Justice. (2021). Advantages and disadvantages of online mediation. *Mediation Matters – Open Justice*. https://www.open.ac.uk/open-justice/sites/www.open.ac.uk.open-justice/files/files/open-justice-week/2021/Mediation%20Matters%20-%20Advantages%20and%20Disadvantages%20of%20Online%20Mediation.pdf
6. Terekhov, V. (2019). Online mediation: A game changer or much ado about nothing? *Access to Justice in Eastern Europe*, 3(2), 33–50. https://ajee-journal.com/online-mediation-a-game-changer-or-much-ado-about-nothing