# Understanding the Voluntary Moderation Practices in Live Streaming Communities

**Jie Cai**
jc926@njit.edu
New Jersey Institute of Technology
Newark, NJ

## ABSTRACT

The moderation task in the live streaming community is challenging due to the interactivity and ephemerality of live text-based communication in the chat. Moderators have to make decisions with time constraints and less instruction, experiencing information overload and emotional toll. I aim to understand their decision-making process, identify the challenges during the process, explore the relationship with other stakeholders in the community. I apply mixed methods (interview, survey, observation) to explore these issues. My dissertation focuses on content moderation in interactive media, and the results can potentially provide guidance to content moderation in interactive media platforms with high interactivity and synchronicity.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

## KEYWORDS

content moderation, volunteer moderators, live streams, moderation strategies, profiling, conflict management

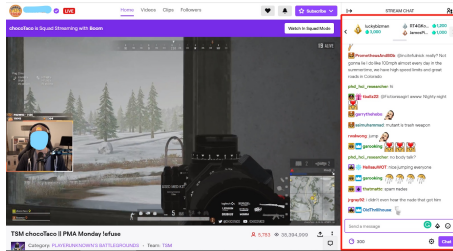| Advisor: | Donghee Yvette Wohn |
| --- | --- |
| PhD Status: | 4th Year |
| Completion Date: | May 2022 |



**Figure 1: Twitch Interface: the Content Producer is Streaming Content on the Left Side of the Screen; Viewers are Commenting in the Highlighted Chatroom on the Right S'ide.**
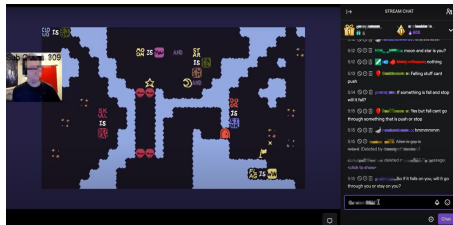


**Figure 2: Twitch Interface from Moderators' View (In the Chatroom: Shortcuts of "Ban""Timeout""Delete" are visible next to the Usernames)**

## MOTIVATION AND PROBLEM STATEMENT

Freedom of speech caused lots of negative, illegal, and harmful online content. In order to regulate the negative content and maintain the civic discourse, a massive workforce of people moderates behind the scenes with the assistance of algorithm and moderation tools. With the adoption of new technology and the evolvement of communities, the moderation practice faces new challenges [10].

Live streaming, as an interactive social medium with high-fidelity computer graphics and video and low fidelity text-based communication, enables the streamer to share the rich ephemeral experience with informal social interaction with viewers in the chat (Figure 1) [6]. The high interactivity of the text-based community prompts a large volume of messages dynamically flowing in the chat and disappearing quickly. Moderators need immediate attention to these messages; high concentration with the time constraint of this kind of information caused information overload and emotion toll [12]. The synchronicity and ephemerality of live streams render different challenges for the hidden labor of volunteer moderators (Figure 2). Thus, understanding content moderation practices in interactive media is crucial for us to identify the challenges they face and provide possible design interventions.

The overall objective of my research is to understand volunteer moderators' relationships with viewers and the streamer and to identify the challenges they face during the moderation process and recommend possible social and technical interventions to increase moderation efficiency and maintain the community with less punitive and more accurate moderation. In order to achieve my objectives, I ask the following high-level research questions:

- What are the strategies that volunteer moderators use in live streaming communities?
- How do volunteer moderators profile violators in live streaming communities?
- What do volunteer moderators manage conflict with the streamer in live streaming communities?

## PRELIMINARY RESULTS AND WORK IN PROGRESS

### Preliminary Results

To answer the first research question, we interviewed 21 volunteer moderators on Twitch. We mapped out 13 moderation strategies and presented them in relation to the bad act, enabling us to categorize from proactive and reactive perspectives and identify communicative and technical interventions (as shown in Figure 3). We identified the flow of decision-making that takes place during the moderation
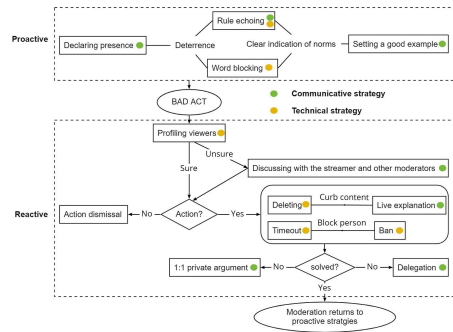
**Figure 3: Moderation Strategies For Before and After a Bad Act Happens. Lines Indicate Relationships, Arrows Indicate Sequence.**
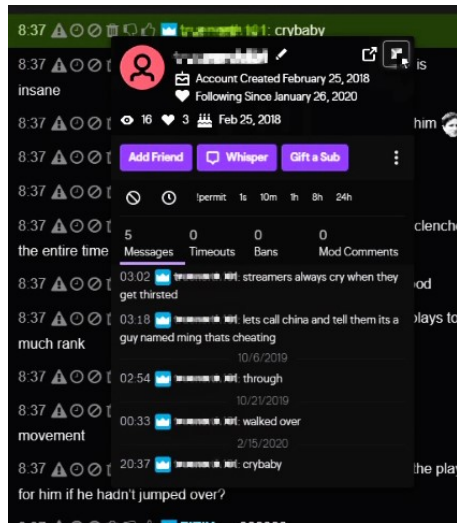


**Figure 4: A Moderator Checks a User's Message History as Part of the Profiling**

process. These practices of volunteer moderation bear similarities but also distinct differences compared with other user-governed communities. The interactivity and synchronicity of live streaming reveal the visible and performative work of volunteer moderation. Instead of considering moderation as blocking content or violators with the assistance of technical agencies, we may also want to take social dynamics into the moderation process and highlight the significance of communicative strategies performed by the human moderator at both the proactive and reactive levels. The affordances of live streaming allow graduated moderation and amplify violators' voices in the moderation process, showing moderators' great effort to increase legitimacy and maintain community members [4].

I also quantified the main strategies and asked user's perceptions of effective moderation strategies in live streaming communities [3]. To supplement these strategies, I also categorized current moderation tools and highlighted the opportunities and the desired new features for bot design in live streaming communities [2].

### Work in Progress

In the second phase, I have developed a new interview protocol based on some of the findings from the first phase and plan to gain more insight into profiling strategy. In addition to a semi-structured interview, I decided to add an observation to consider the specific context of moderation. I asked the moderators to record their moderation screen for an hour and send it to us for a review first. I watched the video and then scheduled the semi-structured interview.

I aimed to understand how volunteer moderators on Twitch create profiles of violators before they decide on what action they will take with the violator. I used criminal profiling as a lens [7]. I found that profiling improved moderators' understanding of violators, and they engaged in complex practices of evidence collection and documentation to create these profiles (Figure 4). These practices happened not just within one community but across different Twitch communities as well as on different platforms. Generally, instead of sanctioning violators, mods preferred to go the extra mile to integrate the violators into the communities. Though they had to sanction some violators, the profiling made the sanction different. We also found that mods across different micro-communities collaboratively worked on violator profiling because of the limited information in the user's profile and limited technical support from the platform (Under Review).

### FUTURE WORK

Volunteer moderators on live streaming platforms often have to work in teams with other moderators and the streamer whose channel they are working for. As each micro-community attempts to set its guidelines, it is common for mods and the streamer to disagree on handling various situations. In the third phase, I plan to apply a mixed method with survey data collection targeting moderators in live streaming communities to statistically test the relationship among the perceived conflict (task,

relationship, process, and normative) [5, 8, 9], moderators' commitment to the micro-community (affective, normative, continuance) [1], and the conflict management styles (integrating, avoiding, dominating, obliging, and compromising) [11]. In the survey, I also asked open-ended questions about conflict incidents and how moderators handle these conflicts to gain context. This research explores the triangle relationships among conflict types in the moderation team, moderators' conflict management styles, and moderators' commitment to the micro-community (specifically, commitment to the streamer or commitment to the viewers). We contribute to understanding conflict management during the moderation process in user-governed online communities and providing insights to micro-community leaders and mods who seek to handle conflicts effectively to grow the micro-community.

## CONTRIBUTIONS

Through three-phase studies to understand the moderators work with both the streamer and viewers in live streaming communities. I highlighted the moderation challenges caused by the affordances of new technology and showed moderators' understanding of and relationship with stakeholders in the micro-community. This work shows the potential to guide community moderation and maintenance in new forms of social media with high interactivity and synchronicity.

## REFERENCES

[1] Patrick J Bateman, Peter H Gray, and Brian S Butler. 2011. The impact of community commitment on participation in online communities. *Information Systems Research* 22, 4 (2011), 841–854. https://doi.org/10.1287/isre.1090.0265

[2] Jie Cai and Donghee Yvette Wohn. 2019. Categorizing Live Streaming Moderation Tools : An Analysis of Twitch. *International Journal of Interactive Communication Systems and Technologies* 9, 2 (2019), 36–50.

[3] Jie Cai and Donghee Yvete Wohn. 2019. What are effective strategies of handling harassment on twitch? Users' perspectives. In *Companion of the ACM Conference on Computer Supported Cooperative Work*. 166–170. https://doi.org/10.1145/3311957.3359478

[4] Jie Cai, Donghee Yvette Wohn, and Mashael Yousef Almoqbel. 2021. Moderation Visibility: Mapping the Strategies of Volunteer Moderators in Live Streaming Micro Communities. In *Proceedings of the ACM International Conference on Interactive Media Experiences*. https://www.researchgate.net/publication/350740235_Moderation_Visibility_Mapping_the_Strategies_of_Volunteer_Moderators_in_Live_Streaming_Micro_Communities

[5] Jason J. Dahling and Melissa B. Gutworth. 2017. Loyal rebels? A test of the normative conflict model of constructive deviance. *Journal of Organizational Behavior* 38, 8 (10 2017), 1167–1182. https://doi.org/10.1002/job.2194

[6] William A Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1315–1324. https://doi.org/10.1145/2556288.2557048

[7] Scotia J. Hicks and Bruce D. Sales. 2007. A Scientific Model of Profiling. In *Criminal profiling: Developing an effective science and practice*. 207–230. https://doi.org/10.1037/11428-011

[8] Karen A. Jehn. 1995. A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly* 40, 2 (1995), 256–282. https://doi.org/10.1177/027507408101500109

[9] Karen A Jehn and Elizabeth A Mannix. 2001. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal* 44, 2 (2001), 238–251. https://doi.org/10.2307/3069453

[10] Jialun Jiang, Charles Kiene, Skyler Middler, Jed R Brubaker, and Casey Fiesler. 2019. Moderation challenges in voice-based online communities on discord. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3. 23. https://doi.org/10.1145/3359157

[11] M Afzalur Rahim. 1983. A measure of styles of handling interpersonal conflict. *The Academy of Management journal* 26, 2 (1983), 368–376. https://doi.org/10.2307/255985

[12] Donghee Yvette Wohn. 2019. Volunteer moderators in Twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. In *Proceedings of 2019 ACM Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3290605.3300390