Categorizing Live Streaming Moderation Tools: 
An Analysis of Twitch

Jie Cai, New Jersey Institute of Technology, Newark, USA
Donghee Yvette Wohn, New Jersey Institute of Technology, Newark, USA

ABSTRACT

Twitch is one of the largest live streaming platforms and is unique from other social media in that it supports synchronous interaction and enables users to engage in moderation of the content through varied technical tools, which include auto-moderation tools provided by Twitch, third-party applications, and home-brew apps. The authors interviewed 21 moderators on Twitch and categorized the current features of real-time moderation tools they are using into four functions (chat control, content control, viewer control, settings control) and explored some new features of tools that they wish to own (e.g., grouping chat by languages, pop out window to hold messages, chat slow down, a set of buttons with pre-written/pre-message content, viewer activity tracking, all in one). Design implications provide suggestions for chatbots and algorithm design and development.

KEYWORDS

Bot Design, Content Moderation, Human-Computer Interaction (HCI), Moderators, Online Community, Twitch Chatbot, Twitch Mod

INTRODUCTION

Live streaming is a mixed media form (Hamilton, Garretson, & Kerne, 2014) that is different from traditional social media in that it is considered a synchronous media with unique attributes such as simultaneity (Scheibe, Fietkiewicz, & Stock, 2016) and authenticity (Tang, Venolia, & Inkpen, 2016) and allows users (broadcasters and viewers) to interact with each other in real time through live video and chat (Wohn, Freeman, & McLaughlin, 2018). The live streaming platform, Twitch, is one of the leading live streaming video service providers that originally focused on games but is increasingly extending to creative content and mobile broadcasting. As of September 2018, Twitch had numerous content categories including IRL (in real life), Creative, Food & Drink, and Travel & Outdoors (Roger, 2018). The streamers are content creators and broadcasters of gameplay or other categories; viewers watch the streaming video then send messages to the streamer or other viewers in a chat interface that is adjacent to the streaming video.

DOI: 10.4018/IJICST.2019070103

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
The popularity of live streams and the success of Twitch have made it a growing subject of academic attention. Most current research on live streams, however, focuses on streamers and viewers, such as streamer or viewer motives (Cai & Wohn, 2019; Cai, Wohn, Mittal, & Sureshbabu, 2018; Friedländer, 2017; Scheibe et al., 2016) and streamer-viewer interactions (Lu, Xia, Heo, & Wigdor, 2018; Wohn et al., 2018), with less but growing attention on the prominent but hidden role of human moderators (Seering, Wang, Yoon, & Kaufman, 2019; Wohn, 2019).

Prior research defines content moderation as “the organized practice of screening user-generated content posted to internet sites, social media, and other online outlets, in order to determine the appropriateness of the content for a given site, locality, or jurisdiction” (Roberts, 2017). Generally, moderators perceive their roles as “filter, firefighter, discussion leader, and content expert” and they moderate content to guide the discussion and to keep down “flames” (Berge & Collins, 2000). Commercial content moderators, who are paid workers, curate content and guard against violations such as racism, homophobic slurs, pornography, and violence (Roberts, 2016). These commercial content moderators usually review inappropriate content that has been flagged by users or detection algorithms. Twitch has commercial content moderators but also enables streamers to appoint their own moderators (also known as “mods”). Mods voluntarily assist the streamer in managing the chat content and are usually unpaid (Wohn, 2019).

Due to the synchronicity of live streams, all the messages are flowing in the chatroom in real time, posing different challenges compared to asynchronous communities such as Wikipedia and Reddit, which also rely largely on volunteer moderators. Technical interventions can, to some extent, reduce the human moderation load, especially in large and fast-moving chats (AnyKey, 2016). Many online communities, such as Reddit and Twitch, apply bots (software robots) to assist the mods in doing moderation practice (Seering, Wang et al., 2019). Current research about using bots for content moderation mainly focus on asynchronous communities such as Reddit (Gilbert, 2013; Long et al., 2017) and Wikipedia (Clément & Guittton, 2015; Müller-Birn, Dobusch, & Herbsleb, 2013), with limited research about bots for moderation on Twitch (Seering, Luria, Kaufman, & Hammer, 2019). Better understanding the moderation tools that mods use every day would help improve the current tool design, reduce the working load of mods, and further benefit the community. The goal of this research is to analyze the features of moderation tools on Twitch into categories that could be generalizable to all other moderation tools and to provide some implications for future tool design.

BACKGROUND

Online Community Moderation and Moderators

One of the primary reasons that online spaces need moderation is because of the negativity that persists regardless of platform. Racism, sexism, and many other prejudices flourish online (Chadwick, 2006) while trolling, flaming, spamming, and flooding messages can disrupt users’ experience (Lampe, Zube, Lee, Park, & Johnston, 2014; Pfaffenberg, 2011). Seven key risk categories need to be addressed by moderation techniques for user-generated content: offensive content, spam, soft hacking, etiquette breach, editorial conflict, copyrighted material, and personal exposure (Coutinho & Jose, 2017).

Communities can use formal and informal methods to enforce standards of appropriate behaviors such as explicit rules, reputation systems, and algorithms (Anis, Börrnert, Rümmeli, & Kuntscher, 2013). Some research suggests that new IT technologies such as moderation systems could deal with information overload and improve participants’ civility (Mironov, Faizliev, Sidorov, & Gudkov, 2016; Resnick, 2002). Researchers have investigated tools to mitigate information overload and manage negativity. Crowdsourcing uses the ability of a large group of people to provide feedback about a single piece of information—for instance, by flagging inappropriate content that is then overviewed by human moderators (Resnick, 2002). Crowdsourcing is commonly employed through rating systems and
Second Life in the Psychology Classroom: Teaching and Research Possibilities
[www.igi-global.com/article/second-life-in-the-psychology-classroom/84814?camid=4v1a](www.igi-global.com/article/second-life-in-the-psychology-classroom/84814?camid=4v1a)

Predicting Medical Resources Required to be Dispatched After Earthquake and Flood, Using Historical Data and Machine Learning Techniques: The COncORDE Emergency Medical Service Use Case