# Toxicity in Twitch Live Stream Chats: Towards Understanding the Impact of Gender, Size of Community and Game Genre

Lukas Dreier
*Graz University of Technology*
Graz, Austria

Johanna Pirker
*Ludwig-Maximilians-Universität, Germany*
*& Graz University of Technology, Austria*

[1]

***Abstract*—Twitch is a leading global live-streaming platform primarily focused on gaming content. However, it is evolving significantly beyond gaming, making it increasingly relevant as a social media and communication platform. Through its chat feature, thousands of streamers interact with their followers in real time. Thereby, the viewers interact with each other or the streamer through a chat. Also here, addressing toxicity and negative posts in the chat, a common challenge on social media platforms, is crucial. Twitch's fast-growing user base creates a potential breeding ground for toxic and hateful behavior. In this first study, we examine Twitch chats to better understand potential toxic behavior. By selecting a diverse group of streamers based on followers, streaming content, and gender, we offer insights into this under-represented research topic and propose ideas for preventing toxicity in live chats.**

***Index Terms*—toxicity, video games, game analytics**

## I. Introduction

With the rise in popularity of live-streaming, addressing negative behavior among chat participants and streamers has become a pressing concern. As social media platforms expand, the presence of trolls, bullies, and toxic individuals also increases. Understanding the triggers and patterns behind such toxic behavior is crucial, yet it remains an underexplored area. Research in this field can contribute to the development of preventive measures like chatbots and content filters. This study focuses on Twitch.tv, a successful live-streaming platform known for its vibrant and interactive community. Our objective is to analyze the live chat messages of diverse streamers to gain insights into toxic behavior during live streams. We employ a comprehensive approach involving logging, parsing, categorizing, and analyzing all exchanged messages to identify and characterize patterns of toxic behavior. Our specific objectives include:

- Investigate gender-related differences in toxic behavior, particularly between female and male streamers
- Explore whether the presence and amount of toxic behavior depend on the reach of a streamer (follower count)
- Analyze the effects of different game genres (e.g., shooter, role-play, strategy games) on toxic behavior
- Identify potential triggers and targets of toxic behavior

Our research findings have the potential to contribute to the development of interventions aimed at preventing and mitigating toxic behavior in live-streaming chats.

## II. Background and Related Work

**Twitch.tv** (TTV) is a prominent live-streaming platform that originated from the website Justin.tv in 2011. Initially focused on gaming content, Twitch grew steadily over the years, expanding beyond gaming to various other areas such as online lectures and tuition [1]. As a result, it has become also an intriguing subject for research. In 2014, Hamilton et al. [2] highlighted the platform's popularity and its ability to foster participation and social communities. They found that viewers are not only attracted to unique content but also to social interaction and participation in the chat with other community members. This emphasizes the significance of the live-stream chat feature. In the Twitch live stream chat, users can react to the stream and other messages by posting their own remarks. These messages typically consist of plain text, references to other chat participants using the '@' symbol, and Twitch-specific emoticons that express feelings and emotions. The extensive functionality of the live stream chat and the resulting opportunities for social interaction contribute to the prolonged engagement of viewers on Twitch [3].

**Toxicity on Social Media Platforms.** Toxic behavior is pervasive across various online platforms, including social media, comment sections, blogs, in-game chats, and live stream chats [4], [5]. It can manifest as aggression, harassment, and bullying, disrupting online communities [6]. Detecting toxic communication is challenging as it can be fluent, grammatical, and hard to distinguish from sarcasm [7], [8]. Many platforms struggle to recognize or delete toxic content. In live stream chats, streamers can address toxicity through automatic detection and deletion of abusive language or by employing moderators who manually remove inappropriate messages and enforce chat rules. Moderators try their best to exclude people with disruptive behavior from the chat, but it remains quite

[1]Preprint: L. Dreier and J. Pirker, "Toxicity in Twitch Live Stream Chats: Towards Understanding the Impact of Gender, Size of Community and Game Genre," 2023 IEEE Conference on Games (CoG), Boston, MA, USA, 2023, pp. 1-4, doi: 10.1109/CoG57401.2023.10333159.

TABLE I
SIZE CATEGORIES OF TWITCH CHANNELS

| Category Label | Number of Followers |
| --- | --- |
| SMALL | $\leq$ 5,000 |
| AVG | $\leq$ 25,000 |
| BIG | $\leq$ 1,000,000 |
| XXL | $>$ 1,000,000 |

challenging to detect these "steadily toxic users" (defined in [9]). Preventive measures are vital as users are susceptible to adopting the behavior of others [10]. However, identifying persistent toxic users remains a challenge. Understanding and researching toxicity on the Internet, especially on social media platforms, is increasingly important given the significant growth in online interactions worldwide.

## III. METHODOLOGY

### A. Dataset

Our research aims to gain a deeper understanding of Twitch chats by analyzing positive, negative, and potentially toxic messages. The dataset comprises chat messages posted during a stream. As publicly accessible chat logs were not available, we developed a Python script to connect to the Twitch IRC interface using sockets. This allowed us to log chats from selected Twitch channels simultaneously. The chat messages were collected over a five-hour period, with individual streams varying in duration (typically 4-5 hours). In total, we logged approximately 100,000 messages, averaging around 3,000 messages per streamer. In the next subsection, we discuss the criteria used to select the 36 Twitch channels we observed.

*1) Selection Criteria:* To determine which streamers to focus on, we considered the following selection criteria: channel size (number of followers), gender, and streamed content. To gather an overview of potential channels, we utilized reputable websites such as TwitchTracker and SullyGnome. These platforms provide comprehensive statistics on Twitch, including average live viewers, total followers, and other relevant information for each channel. To ensure comparability, we introduced size categories (refer to Table I) and assigned them to the selected channels. This approach enabled us to collect data under similar conditions. Our selection resulted in a diverse group of Twitch streamers, with an equal representation of both female and male streamers. The channels were also evenly distributed across the size categories, with the majority falling into the "BIG" category (15 channels) and a smaller number in the "SMALL" category (4 channels). This distribution accounts for the higher volume of messages to analyze generated by larger viewership.

*2) Data Pre-Processing:* To facilitate data analysis and draw meaningful conclusions, the raw IRC messages we logged underwent preprocessing. The initial data was cluttered and challenging to interpret. In the first preprocessing step, we extracted the essential attributes, including timestamp, username, and message content, and organized them into a structured data frame. Additionally, we introduced a categorization attribute to identify the type of each message entry. This categorization distinguished Twitch emotes, spam-like messages (e.g., repeated expressions within a single message), and messages containing references to other users. These preprocessing steps helped to streamline the data and prepare it for further analysis.

### B. Data Analysis

Our analysis methodology primarily relied on Natural Language Processing (NLP) techniques due to the nature of our chat message data. To classify the sentiment of the messages, we employed sentiment analysis, which determines whether the emotion conveyed is positive, negative, or neutral [11].

For sentiment analysis, we used the VADER[2] (Valence Aware Dictionary and Sentiment Reasoner) model, which is specifically designed and trained for analyzing sentiment in social media texts, including short messages like chat messages. VADER provides a polarity score, representing the ratios of positive, negative, and neutral sentiments, which sum up to one. Notably, VADER has shown comparable, or even superior performance compared to other highly regarded sentiment analysis tools [12]. Its suitability for our chat message analysis makes it a valuable choice in our methodology.

To identify which messages are part of toxic communication, we made use of the open-source classifier model Detoxify [13], which is trained especially for toxic comment classification. Similarly to VADER, it outputs a score describing how toxic the input message is. Scores are assigned to labels such as *toxic*, *severe_toxic*, *obscene*, *threat*, *insult* and *identity_hate*. Our analysis utilized a combination of the toxic comment classifier, sentiment analysis, and a manual examination of chat messages within complex contexts to gain a deeper understanding of the chat "mood." The collected data underwent a normality check using the Shapiro-Wilk test. Subsequently, depending on the normality result, significance testing was performed using either one-way ANOVA (for normally distributed data) or Kruskal-Wallis (for non-normally distributed data). Both tests employed a significance threshold of 0.05 to evaluate the null hypothesis that all means are equal.

## IV. FINDINGS

To compare the frequency of positive and negative messages in our dataset, we calculated the average percentage of positive and negative messages across observed live streams. Additionally, we analyzed the proportion of negative messages classified as toxic, as not all negative messages are inherently toxic. This toxic proportion was determined by calculating the ratio of messages classified as toxic to the total number of negative messages. These metrics allowed us to gain a general understanding of sentiment and the prevalence of toxic behavior in Twitch's live stream chat. In the following subsections, we explored the relationships between game genres, game modes, streamer characteristics, and the occurrence of positive,

[2]https://www.github.com/cjhutto/vaderSentiment

TABLE II
GENDER COMPARISON: AVERAGE-% PER GROUP

| Gender | Pos. Messages | Neg. Messages | Toxic*-% |
|---|---|---|---|
| Female | 25% | 10.60% | 22% |
| Male | 23% | 11.80% | **28%** |

* Analysis of variance showed significance with $p < 0.05$

negative, and toxic messages. Our findings provide insights into the dynamics of communication in Twitch's live stream chat, including the factors influencing positive and negative sentiment, as well as toxic behavior. This knowledge can be valuable in developing strategies to prevent and mitigate toxic behavior within online communities.

### A. Gender Comparison

We aimed to investigate whether the gender of the streamer influenced the occurrence of toxic behavior. Streamers' self-identification was used to differentiate between genders. Based on our logged and classified data, we observed that positive and negative messages were equally prevalent regardless of the streamer's gender. However, Table II indicates that male-hosted streams had a six percent higher proportion of negative messages classified as toxic compared to female-hosted streams. This difference was found to be significant in the analyzed streams, suggesting a higher likelihood of toxic communication in male-hosted streams. Our findings provide evidence that gender may play a role in the occurrence of toxic behavior in Twitch's chat, with male streamers experiencing a higher proportion of toxic messages. These results underscore the importance of targeted interventions to prevent and mitigate toxic behavior.

### B. Community Size Dependence

We explore the relationship between community size and the occurrence of toxic communication in Twitch's live stream chat. Streamers were categorized based on the number of followers (as shown in Table I), and average percentage values were calculated, as previously done. The findings (presented in Table III) reveal that smaller channels (¡=5000 followers) have a higher percentage of positive messages, averaging 34%. However, the percentage of negative messages does not significantly vary based on Twitch channel size. Although a slight increase is observed in slightly larger channels, it decreases as the community size further grows. Notably, there is a clear upward trend of toxic behavior with an increasing follower count. Starting at 20% in small streamer communities, the average number of toxic messages notably increases in larger fan bases, particularly within the communities of *XXL* streamers. Overall, our results suggest that as community size grows, the likelihood of engaging in toxic communication also increases. These findings hold important implications for developing strategies and tools aimed at preventing and mitigating toxic behavior in Twitch's live stream chat, particularly within larger communities. Further research is necessary to understand the underlying factors contributing to the observed increase in toxic communication and to devise targeted interventions that effectively address this issue.

TABLE III
COMMUNITY SIZE DEPENDENCE: AVERAGE-% PER GROUP

| Size Category | Pos. Messages* | Neg. Messages | Toxic-% |
|---|---|---|---|
| SMALL | **34%** | 11.40% | 20% |
| AVG | 22% | 10.40% | 21% |
| BIG | 23% | 12.50% | 25% |
| XXL | 23% | 9.40% | 33% |

* Analysis of variance showed significance with $p < 0.05$

TABLE IV
SINGLE-PLAYER / MULTIPLAYER: AVERAGE-% PER GROUP

| Game Mode | Pos. Messages | Neg. Messages | Toxic-% |
|---|---|---|---|
| Single-player | 26% | 12.60% | 20.50% |
| Multiplayer | 23% | 10.70% | 26.60% |

### C. Effects of Different Game Genres

We conducted an analysis to examine the impact of game genre on toxicity within Twitch's live stream chat. We compared single-player and multiplayer games and further explored three popular game genres: Shooter, Strategy, and Role-play/Open World (OW), which were combined as one category. Our findings, presented in Table IV, indicated that streamers playing single-player games received a higher average percentage of both positive and negative messages. Conversely, viewers of multiplayer games experienced a significantly higher rate of toxic messages. In the second comparison (Table V), we observed that all three game genres had similar levels of positive messages, with the strategy genre slightly surpassing the others at nearly 14% for negative messages. The increased occurrence of toxic messages in shooter games could be attributed to the more frequent use of violence within these games. Further exploration and discussion of the potential reasons behind these findings will be presented in the next subsection. These results highlight the importance of considering game genre when examining toxic behavior within Twitch's live stream chat. It provides insights into the potential influence of game characteristics on the occurrence of toxicity.

### D. Trigger and Targets of Toxic Behavior

We now turn our attention to identifying the triggers and targets of toxic behavior. By analyzing the classified negative

TABLE V
DIFFERENT GAME GENRES: AVERAGE-% PER GROUP

| Game Genre | Pos. Messages | Neg. Messages | Toxic-% |
|---|---|---|---|
| Roleplay / OW | 24% | 11.30% | 24% |
| Shooter | 23% | 10% | 27.50% |
| Strategy | 25% | 13.50% | 21% |

and toxic messages, we gained a deeper understanding of the contexts in which these messages were posted. One common trigger we observed was spam postings, where individuals repeatedly requested or begged for something, leading to toxic responses from other viewers. In multiplayer online games, opponents and enemies frequently triggered toxicity, with accusations of cheating or labeling the streamer as weak or incompetent. This aligns with our earlier finding that multiplayer games as streaming content tend to have higher levels of toxic messages compared to single-player games. Additionally, toxic behavior was often directed toward game developers and publishers, addressing issues like delayed updates and unbalanced game mechanics. While there may be other triggers and targets for toxic behavior, our analysis focused on those identified in our dataset. Understanding these triggers and targets provides valuable insights into the dynamics of toxic behavior in Twitch's live stream chat.

## V. Discussion

This paper provides an initial overview of toxic behavior in live-streaming services and its potential triggers. Through the analysis of over 100,000 live stream chat messages, we discovered that the gender of the stream host, community size, and game genre have distinct effects on the occurrence of toxic behavior. We also identified triggers and targets of toxicity. The findings raise important research questions, emphasizing the need for further investigation to enhance our understanding of this complex issue. This knowledge can inform the development of preventive measures and countermeasures to mitigate the impact of toxic behavior. Based on the findings, here are some design takeaways for live-streaming services and game developers. *Moderation Tools*: Implement effective real-time moderation tools to detect and remove toxic behavior promptly. *Guidelines and Education*: Provide comprehensive guidelines and educational resources to users, promoting appropriate behavior and conduct in the chat. *Game Design*: Consider the impact of game genre on toxicity and incorporate mechanics that discourage toxic behavior. *Triggers*: Be mindful of potential triggers for toxic behavior and design systems to prevent or minimize them. *Customization*: Offer users options to customize their chat experience, such as the ability to block or mute specific users.

**Summary of findings**: In this study, we provide a dataset of chat logs from various streamers and their chat activities. Our initial findings indicate that gender, community size, and game genre influence the occurrence of toxic behavior in the chat. Male streamers and larger communities tend to have more toxic interactions. Additionally, multiplayer and shooter games exhibit a higher frequency of toxic messages compared to single-player games, strategy games, or role-playing games. We also identify potential triggers, including text-based and in-stream factors.

**Future Work** : Our findings confirm the presence of toxicity on Twitch and indicate that it is triggered by chat interactions, streamers, and in-game content. These initial results underscore the importance of analyzing toxic behavior in chats for intervention purposes, as well as highlighting potential shortcomings in game design and multiplayer setups. Based on the study's results, future research can focus on: Evaluating the effectiveness of interventions like automated moderation tools, chatbots, or increased human moderation in reducing toxic behavior on Twitch. Conducting longitudinal studies to observe trends and changes in toxic behavior over time, providing a comprehensive understanding of its dynamics. Investigating the impact of moderators and identifying best practices for moderating live stream chats. Analyzing the impact of different game genres on toxic behavior, or comparing genres across platforms to understand the platform's influence. Exploring cultural and social factors that contribute to toxic behavior, gaining deeper insights into its underlying causes. By pursuing these research directions, we can enhance our understanding of toxic behavior and develop effective strategies to foster healthier and more inclusive online environments.

## References


[1] J. Pirker, A. Steinmaurer, and A. Karakas, *Beyond Gaming: The Potential of Twitch for Online Learning and Teaching*. New York, NY, USA: Association for Computing Machinery, 2021, p. 74–80. [Online]. Available: https://doi.org/10.1145/3430665.3456324

[2] W. A. Hamilton, O. Garretson, and A. Kerne, "Streaming on twitch: Fostering participatory communities of play within live mixed media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1315–1324. [Online]. Available: https://doi.org/10.1145/2556288.2557048

[3] Z. Hilvert-Bruce, J. T. Neill, M. Sjöblom, and J. Hamari, "Social motivations of live-streaming viewer engagement on twitch," *Computers in Human Behavior*, vol. 84, pp. 58–67, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563218300712

[4] S.-H. Lee and H.-W. Kim, "Why people post benevolent and malicious comments online," *Commun. ACM*, vol. 58, no. 11, p. 74–79, oct 2015. [Online]. Available: https://doi.org/10.1145/2739042

[5] M. Märtens, S. Shen, A. Iosup, and F. Kuipers, "Toxicity detection in multiplayer online games," in *2015 International Workshop on Network and Systems Support for Games (NetGames)*, 2015, pp. 1–6.

[6] J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Anyone can become a troll: Causes of trolling behavior in online discussions," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1217–1230. [Online]. Available: https://doi.org/10.1145/2998181.2998213

[7] A. Sheth, V. L. Shalin, and U. Kursuncu, "Defining and detecting toxicity on social media: context and knowledge are key," *Neurocomputing*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221018087

[8] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, "Abusive language detection in online user content," in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, p. 145–153. [Online]. Available: https://doi.org/10.1145/2872427.2883062

[9] R. Mall, M. Nagpal, J. Salminen, H. Almerekhi, S.-G. Jung, and B. J. Jansen, *Four Types of Toxic People: Characterizing Online Users' Toxicity over Time*. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3419249.3420142

[10] J. Seering, R. Kraut, and L. Dabbish, "Shaping pro and anti-social behavior on twitch through moderation and example-setting," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 111–125. [Online]. Available: https://doi.org/10.1145/2998181.2998277

[11] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 452–455.

[12] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[13] L. Hanu and Unitary team, "Detoxify," Github. https://github.com/unitaryai/detoxify, 2020.