

# Countering Toxicity in Online Competitive Games

DeYou Chou

University of California, Santa Cruz, United States

\* Corresponding Author Email: [dechou@ucsc.edu](mailto:dechou@ucsc.edu)

**Abstract.** Toxicity within online games has been a huge topic that players have expressed concern with. It is an unhealthy situation that will affect players' emotions and playing experience. This paper will, firstly, dive into the definition of toxicity to make clear of what kind of behaviors are considered toxic. Secondly, the different types of toxicity will be introduced, of which the definition and level of impact on players will be examined. With the types being established, case study research with popular first-person shooter game Overwatch will be conducted to grasp the pattern and impact of toxic actions in game. The research has conducted observations while participating game matches to record all instances of toxicity and searched for online forum posts that discuss about toxicity to gather public opinion of said subject. Last but not least, the currently implemented counteracting methods against toxicity will be discussed while sharing the possibility of future improvement of such methods. With this paper, the problem of toxicity in online games will be shed light on and emphasized to raise more attention from players, researchers, and even game developers, therefore making the community clean and positive for gamers.

**Keywords:** Online competitive games, Definition of toxicity, The mode and impact of actions

## 1. Introduction

Online games, especially competitive games, are corroded with toxicity. In the competitive environment that has ranking system, all players want to win to climb up the ladder. To do so, players will have to cooperate with their teammates to optimize the chance of winning. However, sometimes the team isn't performing to the expected level so that it seems unlikely to win. This is where toxicity brewed. Toxic players at this moment will exhibit toxic behaviors to express their frustration. For example, online games have the space for players, either teammates or opposing team players, to communicate and socialize, like in game chat function. However, this communication method is often abused and turned into online toxicity, where players will blame teammates for not performing good enough, or flame opponents for unfair gameplay. These accusations, most of the time, are emotional, unreasonable, and overly negative [1]. Besides verbally insulting other players, toxic players have their way of playing around game mechanics to perform toxic acts, usually by sabotaging other players gameplay to interrupt and annoy targeted players. Toxicity like these does not simply exist in the game itself. These behaviors can even be carried outside of games, where toxic players will dox and witch-hunt the already wounded victims on social platforms, causing even more harm to them.

The important issue about toxicity is that it can affect players in multiple ways. For the targeted players, they will experience frustration and embarrassment, enjoying less while playing, decide to leave the game prematurely, or even become the toxic player themselves. For game companies, toxic game environment is the least they would want. A toxic community can diminish the players intend to continue supporting the game and less players willing to start playing the game, therefore making the company's revenues drop [2]. This is the apparent issue that has to be solved. Therefore, it is essential to discover and discuss a counter plan against toxicity to improve the gaming environment. This paper will define toxicity, identify different types of toxicity, dive deep into the actual toxicity case with case study, and shed light on a couple possible improvements for this issue.

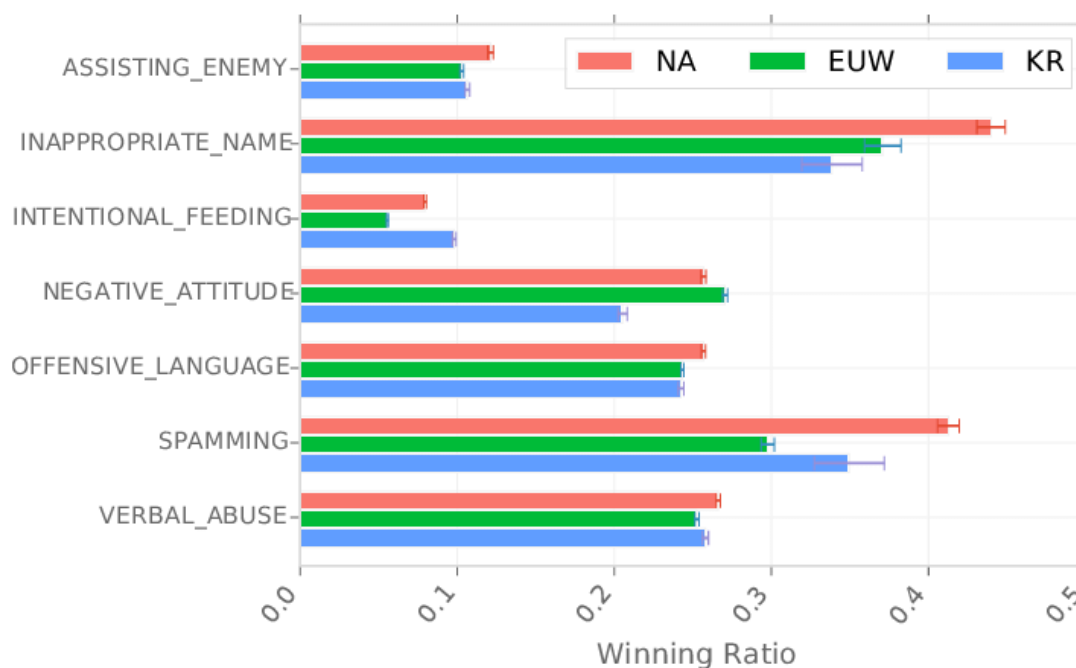
## 2. Define Toxicity

Online games are the popular form of entertainment for players as another style of social life. Many find their voice and confidence through this social environment. However, toxic behaviors

exhibit by certain group of people has polluted the supposedly delightful community to a extend that it can be harmful to other players. Toxicity is the negative behaviors, which have various forms, that players have. These behaviors can occur during the event where the toxic individual is upset when being overperformed by opponent or simply just doing this for fun (in his or her opinion). In most cases, however, toxicity is initiated by one or multiple players criticizing other players for not exhibit satisfactory performance in their own opinion [3]. These abusive behaviors are most commonly seen in competitive game where ranking system is included. Ranking system is a way to show an individual player's power level. It is usually visualized as different tiers, for example, Overwatch, a popular online first-person shooter game, has the rank tier from Bronze to Grandmaster. Usually by winning and performing well in a match, the rank will climb up. With this system, players will strive to win in every ranked match to boost their ranking.

However, sometimes the team is not performing to the satisfactory level, making winning seems impossible. At this moment, toxic players will emerge with various toxic actions to intentionally destroy the game as much as possible as a revenge. Usual types of toxicity include abusive language and communications toward other players, or disruptive gameplay that general members of the community considered to be inappropriate. For example, in Overwatch, intentionally and repeatedly using Mei's ice wall, which is an object that acts like a barricade, to block off team member's attack route and line of sight would be considered toxic by teammates as it is a disruptive behavior to their gameplay experience. The biggest cause for toxicity is the anonymous nature of online gaming alias. One can present inappropriate and harmful behaviors to a much larger extent without suffering the consequence of being judged or punished in real life. Simply put, being toxic does not result in overly severe consequences because no one knows who the abuser is. To clarify, there is no set criteria regarding whether an action is toxic or not. Toxicity is defined by the players, including the enforcer, the victim, and the bystander, from various elements, such as the severity of the action, the corresponding effect that is imposed on the victims, etc. In other words, toxicity should be identified case by case [4]. Because of such nature, toxic players themselves sometimes don't even recognize their actions as toxic, therefore spark more conflict with others, worsening the situation [5].

Toxicity is harmful to other players in several ways. For a team of players that have been matched by the system, toxic behaviors will affect their morale and team coordination, thus resulting in lower chance of claiming victory [4]. For individual players, the targeted player of toxicity will lose the enjoyment of gaming and online socializing. They might also make the decision of leaving the community for good. In the worst case, they will become the toxic players themselves, creating a toxic loop that is detrimental to the community. Fig 1 shows the winning ratio for game matches that include toxic behaviors, and obviously enough, the win rate is noticeably low.



**Figure 1.** Winning Ratio of Game Matches Containing Toxic Actions

### 3. Types of Toxicity

Toxic behaviours refer to the various ways toxic player can negatively affect ally players or opponents. The reasons vary from narcissism (seeking attention) to sadism (enjoying inflicting pain and embarrassment to enemies). Famous MOBA game League of Legends' client has a report system, which categorizes the trolling behaviors as the following types: negative attitude, verbal abuse / hate speech, and intentional feeding / leaving the game / AFK (away from keyboard). Within the following discussion of each type, some of the report types have been combined because of the nature of such actions are identical.

#### 3.1. Negative Attitude

Negative attitude refers to the action in which toxic players express their frustration which is strongly negative that other players will, sometimes unnoticeably, be emotionally affected. Negative attitude will mostly be expressed in team chat or in game voice communication. Generally, negative attitude does not target any specific players, instead, toxic players express their frustration and disappointment through chat for specific reason like attention seeking. Examples for negative attitude includes chats like "You guys' suck, ff (forfeit)" or "gg (good game, sarcastically meaning)" five minutes into match, which affect teammates morale and willingness to win. While negative attitude does not necessary aim at specific players, being inside a negative environment will lower players, teammates and the toxic players themselves, confidence and enjoyment of playing, thus decreasing one's effectiveness in a match and making winning unattainable.

#### 3.2. Verbal Abuse

Verbal abuse and hate speech, which can be called verbal abuse altogether, are more direct and devastating than negative attitude. Verbal abuse is shown with hostile words toward other players. This includes but not limited to flaming teammates for not cooperating, insulting enemies for tunneling (specifically focusing on killing one particular opponent), or telling a player to uninstall the game just because they are new players and don't know how to play the game. The most notable example is "ggez". "Gg", although being a common word being said at the end of match as a civilized and polite farewell for the players, becomes toxic when pairs with "ez", which means easy. This simple four-letters acronym suggests that the team achieves victory with ease, insinuating that the

opposing teams did not have the capability to win, thus further ridicule the already losing team. Hate speech refers to, along with already using strong and offensive language, insulting other players with profanity, racism, and sexism remarks, all of which are capable of damage players' emotion and self-esteem. Sexual toxicity, in particular, is one of the most notorious kinds of hate speech because it is targeted at a specific group, female players. Sexual Toxicity in Gaming, or STIG in short, is the situation where female gamers are being taunted, threatened, and coerced to satisfy imposer's sexual desire, often with the remark of sexism and gender discrimination. Studies have revealed than female streamers on Twitch, a popular online streaming platform, receive eleven times more sexual comments than male streamers [6].

### 3.3. Mechanical Abuse:

Intentional feeding, leaving the game, and AFK, which the researcher categorizes as Mechanic Abuse, are more toward the technical side. Besides using team chat or in game voice communication, using game mechanics to sabotage teammates is another infamous way of toxicity. Intentional feeding describes the action where toxic players will purposefully let opponents defeat them. In almost every online game, this action is detrimental to the team because each kill result in the killer gaining experience points or resources to enhance his or her level, items, or skill trees. In short, toxic players intentionally helping the opposing team because of their own frustration, which can also be connected to the previously mentioned negative attitude. This action can also be known as trolling. Leaving the game, which is self-explanatory enough, is another method of sabotaging the team. Since most online game is symmetrical, meaning that the two competing team have the same number of players, losing one teammate is disastrous that would result in extremely low chance to win. AFK has the same idea with leaving the game. By staying away from the game and not contributing to the team, the team will have lower chance for winning because it is practically the same with leaving the game. Knowing this fact, the team who has AFK players or leavers will result in even lower morale.

Generally speaking, Negative Attitude will have less negative impact on other players, while Verbal Abuse can cause more devastating emotional distress toward others. Mechanical Abuse, although having less emotional offense, greatly disrupts other players' ability to properly play the game, thus results in little to no chance of victory. Although emotionally, Mechanical Abuse doesn't really affect other because it generally doesn't target specific players unless expressing it in chat, at most, it just results in others' disappointment with the unwinnable match. Table 1 shows the level of impact that each type of toxicity brings. Emotional Impact indicates the effect on players on an emotional level, Gameplay Impact refers to the effect on players' ability to play/win the game.

**Table 1.** Types of Toxicity and Their Impacts

|                  | Negative Attitude | Verbal Abuse | Mechanical Abuse |
|------------------|-------------------|--------------|------------------|
| Emotional Impact | Moderate          | Severe       | Normal           |
| Gameplay Impact  | Normal            | Moderate     | Severe           |

## 4. In Game Toxicity and Case Study

To better identify and make example of the toxicity in game, the researcher has conducted a data collection with Overwatch, a notoriously toxic game community. Two methods of data collecting were executed. First, within the 7-days interval, the researcher has closely monitored the in-game chat, voice chat, and match replay. The researcher put focuses on the matches where toxicity occurred and noted down the type of toxicity, frequency, and the consequent actions. Secondly, the researcher searched the official Overwatch forum as well as popular forum site like Reddit for post that specifically mentioned toxicity. The researcher gathered information regarding the toxicity situation, type of toxicity, and the consequent effects with Reddit search tool.

The researcher has personally participated in Overwatch matches to observe toxicity cases. The observation is conducted within 50 matches. 25 of the observed matched are Quick Play matches, which is the normal game mode that has no effect on one's ranking. And to have the competitive

element in play, another 25 observations are conducted in Competitive, where win and lose will affect one's ranking. The research is conducted with covert observation, which means that the subjects of observation were unaware of the research being conducted, thus making the result unbiased [7]. Any cases of toxicity were recorded in a spreadsheet with the following information: Mode, which specifies which game mode does the toxicity occurs, Quick Play or Competitive; Time, which indicates the game time where the action occurred, for example, 2:00 shows that the toxic action happened 2 minutes after the game start; Medium, which indicates the medium used for toxicity actions, the researcher roughly sorted them into three types: Chat Log, Voice Chat, and In-Game Behaviors; Types, which refers to the types of toxicity discussed previously; Description, which provides the description of the toxic situation, toxic speech in chat log and voice chat are quoted and brief explanation of the behaviors are provided; Result, which indicates the win lose status of each match, specific consequences that the toxic actions bring such as the victim leaving the chat room are noted, if any.

Of the 50 matches that are recorded, 11 of them are civilized and friendly, 16 of them have little to no interaction at all, and stunningly, 23 of them have toxic behaviors occurring, making the chance of encountering toxicity 46%, a reasonably high number. Within the 23 matches, 16 cases of Negative Attitude, 15 cases of Verbal Abuse, and 11 cases of Mechanic Abuse were identified and recorded. On top of the disturbingly high number of toxicity cases, 19 of them caused noticeably negative effect on the team, thus proving the severity of toxicity. Notable pattern is also discovered that of the 23 toxicity-infused matches, 16 of them are from competitive mode, which proves that the more competitive the nature of a game is, the more likely toxic actions would appear. Table 2 is the snippet of the recorded matches containing toxicity, with content and types being noted.

**Table 2.** Method 1 Spreadsheet

| Mode | Time | Medium | Type | Description                                           | Result |                                                                                                                                                                                 |
|------|------|--------|------|-------------------------------------------------------|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| QP   | 0:23 | Chat   | NA   | "Lol level diff"                                      | Lose   | Acronym Explained<br>QP = Quick Play<br>Comp = Competitive<br>VC = Voice Chat<br>IGB = In Game Behavior<br>NA = Negative Attitude<br>VA = Verbal Abuse<br>MA = Mechanical Abuse |
|      | 2:45 | Chat   | VA   | "Low level losers, no use"                            | Lose   |                                                                                                                                                                                 |
|      | 3:10 | Chat   | NA   | "Aight imma throw"                                    | Lose   |                                                                                                                                                                                 |
|      | 3:13 | IGB    | MA   | Player starts to AFK                                  | Lose   |                                                                                                                                                                                 |
|      | 4:36 | IGB    | MA   | Player switch to Mei to block entrance with Ice Wall. | Lose   |                                                                                                                                                                                 |
|      | 7:01 | Chat   | VA   | "u guys suck, gg"                                     | Lose   |                                                                                                                                                                                 |
| Comp | 2:21 | Chat   | VA   | "Rein theres a button called W, STUPID"               | Lose   |                                                                                                                                                                                 |
|      | 2:37 | Chat   | VA   | "Wheres my heal then a**hole"                         | Lose   |                                                                                                                                                                                 |
|      | 2:39 | VC     | VA   | "No peels no heals, grow the fuck up"                 | Lose   |                                                                                                                                                                                 |
|      | 2:58 | Chat   | VA   | "healer useless"                                      | Lose   |                                                                                                                                                                                 |
|      | 3:34 | Chat   | NA   | "useless team gg"                                     | Lose   |                                                                                                                                                                                 |
| Comp | 3:02 | Chat   | NA   | "Wheres my heal?"                                     | Win    |                                                                                                                                                                                 |
|      | 3:41 | Chat   | NA   | "u are supposed to heal me"                           | Win    |                                                                                                                                                                                 |
|      | 3:59 | Chat   | VA   | "go kill urself"                                      | Win    |                                                                                                                                                                                 |
|      | 5:11 | Chat   | VA   | "useless Ana, no heal all game"                       | Win    |                                                                                                                                                                                 |
|      | 5:16 | Chat   | VA   | "whats ur worth?"                                     | Win    |                                                                                                                                                                                 |
|      | 5:25 | Chat   | VA   | "healer dont do healer job, trash"                    | Win    |                                                                                                                                                                                 |

The second research is conducted within the Overwatch subreddit on Reddit, which is a famous forum site. The research searched the subreddit using terms like "toxic", "asshole", "bitch" to locate any post that discusses about toxic behaviour [8]. Noted that the time interval is also 7 days, but is conducted in a different time than the first research. The search results are also recorded into separate

spreadsheet with mostly the same content as the first research, minus Time and Medium. A new segment called Toxicity Rate is added to indicate the percentage of people who commented under the post that considered the action to be toxic. Since Reddit is the most used online forum that contains hundreds of post a day, the researcher has only selected 30 posts that discuss toxicity within the time interval. The posts that are asking “Is this toxic?” will be the priority target to be recorded, since Reddit users are more inclined to comment with useful explanation and advice with this kind of posts. Posts that are mostly just rant, unless providing fruitful information, will be more likely to be ignored. Also, because posts with massive numbers of comments will have many comments that provide no useful information, out of topic, or unrelated jokes, the researcher only recorded posts that have less than 100 comments.

Of the 30 posts that are searched and studied, 1 of them belong in the Negative Attitude category, 19 of them can be categorized as Verbal Abuse, and 10 of them fall in the category of Mechanic Abuse. It should be noted that the cases being posted on Reddit were possibly harmful enough that the authors felt the need to post them, so these cases of toxicity would likely to be extreme. Of all the posts that have been searched, 27 of the posts claimed that the initiation of toxic behaviors come from one or several teammates not performing to the toxic player’s so-called satisfaction, showing the pattern that the level of toxicity is correlated to players’ performance [9-10]. The rest 3 posts just describe the trolling situation, where the player is being toxic for comedy purposes or no reason at all. Interestingly 15 of the posts have Toxicity Rate between 50% to 80%, showing there are people on the internet having mixed opinion of what counts as toxic, even if the victims of the toxic actions are clearly negatively impacted. This pattern can be seen as the community member “normalizing” the toxic behaviors as they graduate accept the fact that toxicity exist in online community [5]. Table 3 shows the snippet of the spreadsheet that the researcher recorded, including details like post content and toxicity rating.

**Table 3.** Method 2 Spreadsheet

| Pos<br>t | Mode     | Type | Description                                                                                                                           | Toxic<br>Rate |
|----------|----------|------|---------------------------------------------------------------------------------------------------------------------------------------|---------------|
| 1        | QP       | VA   | OP complains about getting no support while being flamed for no heal as Ana                                                           | 98%           |
| 2        | QP       | VA   | OP was accused to be in the LGBT community and getting the comment "your kind should all be imprisoned" all because of OP's hero pick | 83%           |
| 3        | Com<br>p | MA   | OP's teammate types ggez in chat but uses the spacing to make it looks like its OP's chat                                             | 81%           |
| 4        | N/A      | VA   | Player A insult player B whiile player B "will not listening to you at all"                                                           | 68%           |
| 5        | Com<br>p | MA   | OP was trolled by a fake system message claiming that OP is reported                                                                  | 100%          |
| 6        | Com<br>p | VA   | OP was being insulted with words like "pussy" for playing Bastion                                                                     | 32%           |
| 7        | Com<br>p | VA   | OP got flamed too hard and frequently that he decided to quit the game                                                                | 100%          |
| 8        | QP       | MA   | OP's teammate got tunneled by 4 enemies.                                                                                              | 67%           |

## 5. Currently Implemented Counter Methods

Acknowledging the fact that toxicity in game will lower player’s willingness to keep playing, game developers have implemented several counter methods to relieve the game from corroded with toxicity.

Profanity filter has been implemented in almost every online games nowadays. A list of common profanity words will be collected and conducted in the filter. Should a profanity chat be typed out by a certain player, that words will be detected and censored, normally with asterisk mark. For example, if a sentence “fuck you” is typed out by toxic player, the filter will automatically censor the sentence

into “\*\*\*\* you”. Some games even record the frequency of profanity filter being triggered. Should a player alert the profanity filter numerous times, he or she will be banned from using the chat function, or even from playing the game. This method, although already filtering out the most baleful language, cannot detect strong language without the use of profanity. For instance, a carefully worded insult without using profanity like “KYS (kill yourself)” might not be detected by the filter. While expanding the filter range to less hurtful language might help with keeping the chat clean, it might block important message that players try to communicate with their teammates. Therefore, profanity filter like this, although being affective with extremely strong language, falls short when dealing with less harmful but still effective languages. Nevertheless, it does prove to be valuable in terms of blocking vile language that is too insulting.

Report system is another counter measure for toxicity that is being utilized by most of the game for players who wish to have punishment imposed on certain toxic players [7]. This feature is often visible and selectable at the end of a match. Players who wish to report inappropriate behaviors can simply click the report button, write a few descriptions of the situation, and let the system review the match for validation. Mostly, this feature will be used to reflect toxic behaviors that players observed. Take League of Legends’ report system as example, the automated Disciplinary System will be triggered if someone submit a report. The Disciplinary System will analyze the match data, read the chat log, and progress through other internal verifications. If the report is indeed truthful, a penalty might be placed on the reported player. The player’s game profile, match history, and frequency of being reported will also be taken into account for enforcing a penalty. This review system is all done by computer system, which of course lack accuracy when dealing with the “grey area” toxicity. In this case, a real person’s review might be the most beneficial. However, with thousands of reports being submitted, this method is not achievable. This system provides a way for the players to counter toxicity, but it can be often abused. For example, it is reasonable for a player to report his or her teammate for constant swearing in chat. However, some players like to report a teammate for guiding them how to play. It is of course not a reportable situation as the teammate is teaching them the winning strategy and consider for the whole team, but these players don’t like taking advice, especially if the advice is not agreeable to them, thus consider this action as toxicity and report it. In short, same as profanity filter, many online game’s built-in report system is able to sort out blatant toxic behaviors, especially Mechanical Abuse, and impose appropriate punishment, however, it struggles with identify obscure toxicity actions and users that abuse the report system.

## 6. Future Improvement

Although several counter acts have been implemented to prevent toxic players, quite a few newly invented models have been produced to better help the existing method. For example, researchers from University of Saskatchewan have presented a model that can determine if a game match is toxic by using machine learning technology to detect toxic speech in voice chat. They gathered multiple video streams of Overwatch gameplay and analyzed the audio clip with specific approach such as the timing of speech and the spoken content of voice chat. These data will be parsed into the learning model to produce a binary outcome to determine whether the match is toxic or not. This model, the researchers noted, is an automated system that can be implemented into Overwatch’s existing report system, where it can analyze the match to find toxicity much more accurate than the current implemented automated detection system [1-2].

## 7. Conclusion

After deep study of toxicity in online competitive games, it becomes clear that toxicity can negatively impact players or even the whole gaming community. Toxicity actions vary from expressing negative attitude to being verbally offensive. Commonly seen toxic behaviors include, but not limited to, making sarcastic remarks to express frustration, insulting other players, sabotaging

other's gameplay, and more severe cases like making racist comment or doxing player publicly on social media platform. Competitive games like League of Legends, Overwatch, and Apex: Legends are the most polluted gaming community because of players intend to win and the backfired negative emotion after realization of low winning chance. Although several counter methods are implemented by game developers, such as profanity filter and report system, obscure cases of toxicity still cannot be easily detected with the said methods. In the future, more advanced method of detecting toxicity, like machine learning technology, can be implemented to identify said behaviors so that game developers can impose appropriate punishment, thus remedy the damage that the toxic player inflicted to the community.

## References

- [1] Reid, E., Mandryk, R., Beres, N., Klarkowski, M., & Frommel, J. "Bad vibrations": Sensing toxicity from in-game audio features. 2022, *IEEE Trans. Games*, 1–11.
- [2] Martens, M., Shen, S., Iosup, A., & Kuipers, F. Toxicity detection in multiplayer online games. 2015 *Int. Net. Sys. Sup. Game*. **738**
- [3] de Mesquita Neto, J. A., & Becker, K. Relating conversational topics and toxic behavior effects in a MOBA game. 2018 *Enter. Comput.*, **26**, 10–29.
- [4] Türkay, S., Formosa, J., Adinolf, S., Cuthbert, R., & Altizer, R. See no evil, hear no evil, speak no evil: How collegiate players define, experience and cope with toxicity. 2020 *CHI Conf. Hu. Fac. Comput. Sys*. **38**.
- [5] Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. Don't you know that you're toxic: Normalization of toxicity in online gaming. 2021 *CHI Conf. Hu. Fac. Comput. Sys*. **201**.
- [6] Fredman, L. A. Not just a game: Sexual toxicity in online gaming hurts women. 2018, *Univ. Tex. Aus*.
- [7] Lapolla, M. Tackling toxicity: Identifying and addressing toxic behavior in online video games 2019, *Set. South Or*. **341**
- [8] Sengün, S., Salminen, J., Jung, S.-gyo, Mawhorter, P., & Jansen, B. J. Analyzing hate speech toward players from the MENA in league of legends. 2019 *CHI Conf. Hu. Fac. Comput. Sys*. **319**
- [9] Kwak, H., & Blackburn, J. Linguistic Analysis of Toxic Behavior in an online video game. *Lecture 2019, No. Com. Sci.*, 209–217.
- [10] Neto, J. A., Yokoyama, K. M., & Becker, K. Studying toxic behavior influence and player chat in an online video game. 2017 *Int. Conf. Web Intel*. **83**.