CHAPTER I

INTRODUCTION

1.1 Background

Moderation can be defined as the process of monitoring content or behavior in an environment (Digital Minds BPO, 2024). According to Cecillon et al. (2019), moderation includes identifying and making moves against users who participate in oppressive ways of behaving, like individual assaults or discrimination, to keep a protected and respectful online community whether online or in the real world. Moderation can also be community-driven, where users are encouraged to report violations of platform policies (Gillespie, 2018).

The process of monitoring and managing the conversations that occur on online platforms is referred to as chat moderation. As defined by Oxford Learners Dictionaries, the term ‘chat’ refers to a mode of communication through text in real-time, typically utilizing a device such as a computer, smartphone, or tablet. This form of communication allows users to exchange messages instantaneously, thereby facilitating quick and interactive conversations. Chat moderation plays a crucial role in maintaining the quality and safety of these interactions. Digital communication is basically talking to other people using electronic devices. By sending messages and information via computers, cell phones, and tablets, this let us exchange information with people instantly. According to Gallager (2008), a digital communication system is a communication system that uses digital. This part of digital communication specifically involves the use of computer technology to connect people. As stated by Latvys et al. (2023), digital communication through computer technology has revolutionized the way humans interact, and forming a new type of communication, Computer-Mediated Communication (CMC).

CMC is a term that refers to all forms of communication that occur through the use of computers and digital networks. According to Bekar and Christiansen (2018), computer-mediated communication or CMC is now increasingly used compared to face-to-face interaction, especially in developed countries. A lot of jobs and daily activities are done via computers or mobile devices. Meanwhile, Simpson (2002) said that CMC can be divided into two types: synchronous and asynchronous. If it is synchronous, it means that the chat occurs directly at that time, so people can reply to each other quickly. But if it is asynchronous, there is a time lag between one reply and the next. This research focuses on synchronous communication, specifically the Call of Duty Mobile (CODM) global chat feature. CODM is the mobile edition of the popular game developed by Activision and Tencent Games, and was released in 2019. Battle royale and team deathmatch are the new additions to the game that were previously absent from the previous Call of Duty series. It ultimately boosts the full experience of gaming, as CODM includes a global chat where all CODM players can share information and communicate in real time.

Because in CODM we can chat directly and easily, the possibility of sexually offensive words is increased. Direct interaction does make the game feel more exciting and fun, but on the other hand, it also gives some players the opportunity to say things that make other people uncomfortable or even hurt. This kind of behavior can make the game atmosphere less friendly and make other players uncomfortable.

Therefore, the moderation system is very important so that the gaming world remains a fun and fair place for everyone. Effective moderation can help prevent the spread of sexual harassment. Therefore, moderation is responsible for finding and dealing with bullies, for instance, players that tend to attack or discriminate other players, so the online community

stays a safe and respectful environment (Cecillon et al., 2019). Maintains a sense of comfort and respect in the community. Good moderation also contributes to a more stable and enjoyable online environment for everyone. Here, moderation monitors what people say in the game chat.

1.2 Theoretical Framework

1.2.1 Computer-Mediated Communication

Computer-mediated communication (CMC) is the way people talk or interact with each other using digital devices such as computers, phones, or tablets. It includes things we use every day like social media, messaging apps, email, and online forums (Herring, 2004). CMC can happen in two ways. First, it can be real-time or synchronous, like chatting with someone or having a video call (Simpson, 2002). Second, it can be asynchronous, which means there's a delay, like when sending an email or leaving a comment on a forum (Simpson, 2002) and the other person responds later.

Synchronous communication is communication that occurs directly or in real time between the participants involved. As Lowenthal, P.R. (2022) stated that synchronous communication is communication that happens at the same time or in real time. This communication is the most natural and basic way for humans to connect and exchange information. As technology developed, synchronous communication also evolved with the emergence of digital communications such as telephone, radio, television, and the internet, which in CMC includes various types of text-based online chat, computer, audio, and video conferencing (Simpson, 2002). On the other hand, different from synchronous communication which occurs directly and instantaneously, asynchronous communication is

the exchange of information that occurs without the need for participants to be online simultaneously (Simpson, 2002). Asynchronous CMC includes email, message boards (Suler, 2004), and discussion forums (Herring & Stoerger, 2014).

However, the immediate and interactive nature of synchronous communication can also create opportunities for digital crimes, including sexual harassment and other forms of inappropriate behavior. There are a lot of things that encourage people to commit sexual crimes in digital communication, one of the factors that encourages this to happen is the existence of anonymity or unknown accounts, a concealment of identity (Suler, 2004), and are usually referred to as "anonymous" which means not identified by name (Herring and Stoerger, 2014), which gives them more freedom to commit these crimes. Herring & Stoerger (2014) stated that anonymity reduces social accountability, making it easier for harassers to engage in hostile and aggressive actions. Because, when communicating online, people with an unknown identity (anonymous), their behavior tends to be different than when communicating face to face (Liu, 2017). However, this behavior can lead to disinhibition.

Disinhibition is a phenomenon in which individuals tend to reduce or lose the social inhibitions that normally regulate their behavior. According to Joinson (2007), individuals may lose their typical social inhibitions and act more improperly or violently when they are online. This phenomenon is known as the disinhibition effect. This impact might worsen violent and discriminating actions in the setting of online gaming, especially in CODM, creating a toxic atmosphere. The online disinhibition effect is a phenomenon that explains why online anonymity frequently results in bad actions. The disinhibition effect, as defined by Suler (2004), explains how people are more likely to act in ways that they would normally avoid in person when they are protected by the anonymity of the internet. Many types of

cybercrime, such as sexual harassment and cyber exploitation, have been linked to this impact. (Joinson, 2007).

To mitigate the dangers of anonymity, particularly in preventing sexual crimes, online platforms have created moderation mechanisms, which are methods or procedures that maintain safety, security, and productivity across platforms. Through the prevention and redress of inappropriate or harmful behavior, this process seeks to build a safe, civil, and productive online environment. Moderation mechanisms also shape and maintain most language society on online platforms. There are three types of moderation: Automated, human, and community-driven. Automated moderation is the algorithm and techniques used to identify and filter information according to predetermined standards. According to Cecillon et al. (2019), human moderation means that real people are in charge of checking and managing content that has been flagged. However, community-driven moderation involves users who help by reporting and rating content, which helps keep the community safe and respectful (Gillespie, 2018).

1.2.2 Language Society

Language society is concerned with the relationship between language and society. It looks at how people use language in different social settings and how language and society influence each other. Language helps to shape social relationships and identities because language reflects culture and social systems (Balboni, 2018). Discussing how language works in a community, especially in computer-mediated communication environments, is the focus of language communities. In CMC, language communities are like small versions of real-world social interactions, where the way people speak demonstrates and supports social

rules, values, and power (Herring, 2004).

With the concept of language society, community standards become important in guiding how people speak to each other. These rules help to ensure that the language used conforms to what is accepted by the group as normal and respectable. This helps to create a more respectful and understanding environment for everyone.

1.2.3 Community Standard

Community standards are important rules that guide how people should communicate, especially on digital platforms. These rules are based on social norms and help make sure that conversations stay respectful. In online games, community standards play a big role in building good relationships between players. They clearly explain what behaviors are not allowed, like harassment, hate speech, and the use of offensive or taboo language. By following these rules, players can enjoy a safer and more comfortable gaming experience (Vasalou et al., 2008).

Call of Duty, including its mobile version, also has strict community standards to keep the game respectful and safe for everyone. These rules prohibit all forms of hate speech and abusive language so that all players feel protected. By having clear rules, Call of Duty strives to create a supportive gaming space, where everyone can play comfortably without feeling harassed or discriminated (Call of Duty® Code of Conduct | FPS Game Terms, n.d.).

1.2.4 Taboo Words

Taboo words are words or phrases that are considered rude or inappropriate in certain situations. Taboo words are commonly related to sensitive or controversial topics, such as topics about sexuality, death, or something considered sacred. According to Jay (2009), taboo

words are words or phrases that should not be used or are restricted in society because they can be offensive or considered inappropriate. Taboo words have a significant impact on communication in chat rooms, digital spaces where people interact for a variety of purposes (Camtepe et al., 2004), such as attracting attention, expressing contempt, provocation, or mocking authority.

According to Ningjue (2010), taboo words are divided into five types: cursing, profanity, obscenity, epithet, and sexual harassment. In this study, the focus is sexual harassment. Sexual harassment refers to inappropriate behavior related to sexual matters that can make someone feel uncomfortable or disrespected. The use of taboo words in this context is a serious problem because it involves offensive language that is intended to hurt or demean others. According to Jay (1992), the use of sexual harassment is usually demeaning, humiliating, and an attempt to exert power over someone based on their sex or gender.

1.2.5 Sexual Slurs

Sexual slurs are terms or phrases that demean individuals based on their gender identity, sexual orientation, or other sexual characteristics. According to Bianchi (2014), slurs are offensive expressions that target individuals or groups based on attributes such as race, nationality, religion, gender, or sexual orientation. The use of these terms often reflects patriarchal social norms and gender-based discrimination. In online contexts, particularly on anonymous or semi-anonymous platforms, sexual slurs are frequently used to intimidate, humiliate, or assert dominance. Jay and Janschewitz (2008) have claimed that swearing does not always indicate impoliteness and derogatory, it is usually one's unplanned expression response to the surprising event. They indicate that the derogatory aim of swearing depends

on hearers' interpretation. In the meantime, Bowers and Pleydell-Pearce (2011) have affirmed that swear words as a language form influence conception.

Sexual slurs are terms or phrases used to demean someone based on their gender identity, sexual orientation, or other sexual characteristics. According to Bianchi (2014), slurs are expressions that offend individuals or groups based on factors such as race, nationality, religion, gender, or sexual orientation. The use of these words often reflects patriarchal social norms and gender-based discrimination. In the digital world, especially on anonymous or semi-anonymous platforms, these slurs are often used to intimidate, humiliate, or show power. According to Jay and Janschewitz (2008), slurs are not always meant to be rude or insulting. Sometimes, people say them as a quick reaction to something surprising. They also explained that the meaning of a slur can change depending on how the listener understands or interprets it.

The use of sexual slurs can have a terrible impact on a person's psychological condition. Victims who are the target of this kind of slur often feel anxious, sad, and even lose their self-confidence. In many cases, people who experience bullying, including through sexual slurs can experience serious emotional disorders, even to the point of having thoughts or desires of suicide (Felmlee et al., 2019). The problems do not stop there. If slurs like this continue to be used and are considered normal by society, many people will become insensitive to their impact. They will consider the insults as something normal or funny, even though they are actually very painful. This insensitivity makes sexual insults easier to spread and accept in everyday conversation. As a result, people start to see these insults as normal, and they become part of everyday culture without anyone stopping them. Derogatory language is considered normal, and this reinforces negative views and discrimination against

certain groups (Difranco & Morgan, 2023). Over time, it also influences how people in society behave and what they think is fine, which makes it easier for harassment and slurs to keep happening (Felmlee et al., 2019).

Sexual slurs can be classified into three main categories:

1.2.5.1 Reinforcement of Stereotypes

This refers to words or actions that make people believe false or unfair ideas about a group, which can lead to discrimination. Sexual slurs are often used to insult people who do not fit the usual expectations of how women should look or act (Felmlee et al., 2019). Examples include insults about appearance, like "ugly". The word "ugly" falls into the category of Reinforcement of Stereotypes based on Felmlee et al. (2019) because it is used to insult a person's physical appearance in a way that enforces traditional beauty standards, especially toward women. Felmlee et al. explicitly state that terms like "ugly", along with other insults about appearance (e.g., "fat," "old").

1.2.5.2 Gendered Insults

These are derogatory or discriminatory terms directed towards someone based on their gender. Terms like "bitch", "cunt,", "slut" , and "whore" are highly popular across social media platforms. One study showed that derogatory words of that nature were used over 2.9 million times within a week. This clearly indicates the prevalence of such terms and how damaging they are in nature. (Felmlee et al., 2019).

1.2.5.3 Sexual Orientation Slurs

These are violent words used to abuse a person based on their actual or perceived sexual orientation. Take for instance the derogatory term "fag" It is especially used

to demean people and can incite further hate and prejudice against the members of the LGBTQ+ community (Anderson & Lepore, 2011).

1.3 Review of previous studies

A study by Lim et al. (2024) entitled “Trash-talking versus Toxicity: An Analysis of/All Chat Exchanges between Southeast Asian Players of an Online Competitive Game” analyzed toxic behavior in online competitive games using the /all chat function in Dota 2 in Southeast Asia. It made a distinction between toxic behavior such as cursing and bullying other players or making fun of others not to hurt them, but to make the game more fun and challenging. Out of 26 chat logs studied, instances of both types of communication were present. The results show that bad behavior in chats is rare. On the other hand, trash-talking an opponent is quite common and in line with the competitive spirit of the game. This research analyzes the difference between trash talking and toxicity in video games and creates a new definition while explaining the meaning of both terms. The aim of the research is to deepen the understanding of the online gaming world by analyzing how players interact and how that interaction affects the community and the gaming experience.

A study by Maharani et al. (2024) titled "Understanding Toxicity in Online Gaming: A Focus on Communication-Based Behaviors towards Female Players in Valorant" talks about different types of toxic behavior in online games, especially in Valorant. The researchers used a method called participant observation methods through game recording to see how players interact in competitive mode. The results show that many female players experience verbal harassment, sexist comments, and disrespectful treatment, especially when they use voice chat to communicate during the game. Many of them choose not to speak while playing.

This study also shows that existing moderation systems cannot prevent toxic behavior even if the reports are available to game developers. This study relates to the topic of the research conducted because it discusses the problem of abusive communication and the need for mitigation in the online gaming environment. Focusing on the experiences of female players provides a more accurate perspective on how toxic behavior affects certain groups in the gaming community.

A study by Sengün et al. (2019) titled "Exploring the Relationship Between Game Content and Culture-based Toxicity: A Case Study of League of Legends and MENA Players" examined how a player's culture affects toxic behavior in League of Legends (LoL). The researchers found that certain parts of the game, like certain characters or map settings, can lead to players behaving in a way that is unkind or mean. This often happens when someone makes fun of or insults a player because of their culture. This kind of hate speech is similar to the toxic behavior that we see in many online games. This research shows that some game features can make interactions worse, especially in communities with players from different cultural backgrounds. This is similar to how sexual insults can be an issue in games. Just like the problem with sexual slurs in games, this study shows that some game features can make bad interactions even worse, especially in communities where players come from different cultures. From the analysis of game chat in this study, it appears that players often relate characters in LoL to cultural stereotypes that exist in the real world. This can develop into rude or offensive remarks. In the context of sexual slurs, players frequently target female gamers, using gendered insults in ways that reflect broader societal biases. The findings of this study are in line with the persistent problem of gendered insults in games, where toxic behaviors are often rooted in both cultural and gender stereotypes. This research

suggests that game developers should pay more attention to designing games that can reduce negative behaviors, such as insults or harassment, for every players to have a safer and more comfortable gaming experience.

A study by Kremin (2017) explored the social and emotional functions of gender-directed swearing and slurs in her study, "Sexist Swearing and Slurs: Responses to Gender-Directed Slurs." Kremin discussed how slurs not only reflect societal gender roles but also serve as a mechanism for policing deviations from those roles. Her research highlighted that slurs directed at women tend to target sexual behavior (e.g., "slut," "whore"), appearance (e.g., "ugly"), and emotional behavior (e.g., "bitch"), while slurs directed at men often challenge masculinity by implying weakness or homosexuality (e.g., "faggot," "wimp"). The study proposed that gender-congruent slurs, insults that align with societal stereotypes for each gender, cause stronger emotional responses. Kremin's work also emphasized that slurs carry emotional weight that can be measured by physiological responses such as those captured by EEG, supporting the idea that these expressions are internalized over time due to cultural conditioning.

A study by Jeshion entitled “Slurs and Stereotypes” (2013) discusses the nature of slurs from a philosophical and semantic perspective. In her research, Jeshion criticizes the view that insults always carry stereotypical meanings about certain groups. She argues that although stereotypes often arise when insults are used, the meaning of the word does not always contain the stereotype itself. Instead, Jeshion supports the view that the power of insults comes from the negative emotions conveyed, such as hatred or degrading feelings towards certain groups. This approach emphasizes that the speaker's intention and the social situation greatly influence how much an insult can hurt others. Jeishon's work is relevant to

this research because it provides a theoretical basis for understanding why some insults can still feel offensive, depending on the context in which they are used. In the case of CODM, moderation is conducted by an automated system that is unable to fully understand the speaker's intention or the social meaning of the utterance. Jeshion's theory shows the weaknesses of such a system. The idea helps explain why some insults that seem neutral on the surface can still be hurtful in certain gaming interactions, and why context-based analysis is important to determine how effective moderation is.

A study by Lauren Ashwell' entitled "Gendered Slurs" (2016) explores how gendered slurs such as "slut," "bitch," and "sissy" still function as slurs. Many classic theories of slurs argue that a slur is only a slur if there is a neutral version of the word. However, Ashwell argues that many gendered slurs, such as "slut" or "bitch," do not have a neutral version. Ashwell notes that many classical theories, especially those dealing with racial or ethnic slurs, assume that every slur has a neutral, non-derogatory version (e.g., "black" vs. "nigger"). However, Ashwell challenges this view by highlighting that many gender slurs do not have such a neutral equivalent. For example, the word "slut" not only describes behavior, but also carries with it moral judgments and cultural norms that make it derogatory. Ashwell's findings are relevant to this study, which analyzes the moderation of sexual slurs in CODM. Her argument supports that slurs can be harmful even when not explicitly offensive. This is particularly important in assessing how automated moderation systems respond to words like "slut" or "bitch," which may go undetected by filters despite their derogatory meaning. By challenging the assumption that neutral equivalents are necessary, Ashwell emphasizes the importance of considering context, gender norms, and implicit bias in the design and evaluation of moderation systems.

1.4 Research Questions

Based on the previously identified phenomena. Two questions become the focus of the research, they are:

1. What sexual slurs are moderated and unmoderated in CODM chat?
2. What categories do the sexual slurs found in CODM fall into?

1.5 Objectives of The Study

In conducting this research, the research has two objectives:

1. To identify the moderated and unmoderated sexual slurs in CODM chat.
2. To classify sexual slurs found in CODM into specific categories based on their characteristics.

1.6 Scope of The Analysis

This analysis focuses on why sexual slurs are moderated within the global chat room of CODM. The researcher will collect, filter, identify, and explain the reason why sexual slurs are moderated in the global chat room of CODM. The analysis applies Flemlee and Anderson's theory about the categories of sexual slurs to determine why the words are moderated.