Proceeding Paper

Conscious or Unconscious: The Intention of Hate Speech in Cyberworld—A Conceptual Paper †

Noramira Fatehah Azman [ID] **and Norena Abdul Karim Zamri *** [ID]

Faculty of Communication and Media Studies, Universiti Teknologi MARA (UiTM), Cawangan Melaka Kampus Alor Gajah, KM 26 Jalan Lendu, Alor Gajah, Melaka 78000, Malaysia
* Correspondence: norena@uitm.edu.my
† Presented at the International Academic Symposium of Social Science 2022, Kota Bharu, Malaysia, 3 July 2022.

Abstract: Exponential growth in social media has led to the increasing popularity of hate speech and hate-based propaganda. Hate speech or malicious expression refers to the use of offensive, violent, or offensive language and its religious conduct with a specific group of people who share a common property, such as gender, ethnicity, race, or beliefs. Online hate diffusion has now become a serious problem as it creates a series of international initiatives aimed at defining problems and developing effective countermeasures; this study delves into the exploration of the intention of hate speech posting on social media, especially on Twitter. Both dramaturgical models of social interaction and cultivation theory were used to explain the hate speech culture phenomenon. A qualitative method is proposed for this study as part of the exploration. Results revealed that most of the previous studies on hate speech focused on the field of computer science but rarely on the communication field. The paper presents the results of past studies and shows the new proposed framework. The investigation suggests future directions for the problem and possible solutions; it starts with the background of the research, the explanation of the problem, the meaning of the research, and pursuing the research questions and goals of the research before finally explaining the limits.

Keywords: hate speech; cyberworld; social media; Twitter; offensive language; malicious comment

1. Introduction

There is no denying that Malaysia is one of the countries that is not left behind in pursuing technological modernity. Today, technology is evolving with its variety of sophistication. The traditional details lead to more modern technical pieces of information in line with the current of modernity over time. One of the latest technologies is social media, where we can socialize and communicate well with people worldwide.

Twitter is one of the most popular social media platforms, with almost every adolescent and young adult having an account. Between 2005 and 2010, the popularity of Internet social networking devices such as Twitter skyrocketed. Twitter was founded in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and was launched in July of the same year [1]. According to [2], as of early 2019, Twitter has over 330 million active monthly users. Twitter is unique because of the brevity in which users interact with other users; Twitter messages, or tweets, are limited to 280 characters, a gif, an emoji, a tag location and a poll, we can "tweet" anything we want with our own freedom of speech. Twitter offers three different ways for users to interact with each tweet: a like button, a retweet button, and a reply button. The like button is to show that we love the thing by clicking the love-shape button and our mutual friends on Twitter will get a notification by the apps. Other than that, the Retweet button is a sharing button, so our followers see what we share. The reply button is where users can comment on the tweet to give feedback on some tweets.

According to [3], the numbers posted on Twitter's advertising resources showed that Twitter had 4.4 million users in Malaysia in early 2022; this number indicates that 13.3% of

Malaysia's population was reached via Twitter advertising at the time. However, because Twitter limits the use of the platform to people over the age of 13, it may be helpful to know that 16.7% of Malaysia's "qualified" audiences will use Twitter in 2022. Furthermore, Twitter's Malaysian advertising reach earlier this year accounted for 14.9% of the local Internet user base (regardless of age). Twitter's popularity as a new medium for online communication has caused language issues that require this research, as it clearly imposes word restrictions on each update; this limitation also created another issue with messages that tend to be ambiguous when communicating over Twitter. With only 280 characters, a photo or video, a gif, an emoji, a tag location, and a poll, we can "tweet" anything we want with our own freedom of speech.

Malicious posting or hate speech has frequently been examined in relation to the communication of (extreme) right-wing populist leaders [4]; these players are also becoming increasingly linked to the unchecked spread of misinformation [5]. As it is used by millions of people or we call netizens, some negative things like dissemination of false information, identity fraud and mostly hate speech will occur uncontrollably as social media like Twitter has no gatekeeper; these platforms partially guarantee the user's freedom of expression, so this means it can be used to attack or insult others. Safety concerns result from this circumstance. Online attacks and abuse have been shown to cause not only psychological and physiological health problems for victims, but also self-harm and even suicide [6]. Therefore, our focus is to analyze the intention of hate speech in controlling hate speech in the cyberworld to avoid more negative impacts occurring to users.

In accordance with [7], malicious language or hate speech is used to identify messages that violate existing legal norms and require government regulation or publicly shared statement that incites, encourages, justifies, or are based on a particular group of hatred, discrimination, or hostility. For instance, race or ethnicity, beliefs, abilities or disabilities, gender, age, sexual orientation, or gender identity. Other than that, ref. [8] defined hate speech as showing disgust as it spreads, incites, and promotes hatred, violence, and discrimination against individuals or groups because of their protected nature; these include "race", ethnicity, religion, gender, sexual orientation, and disability, among other social distinctions.

Davidson et al. [9] said that the definition of hate speech is a language of expression that describes the use of words that are hatred or blame, offensive or insulting to the subject. Hate speech as a jargon is no longer an individual-dominated territory in the legal world, but within reach of amateurs. In the context of the Internet and social networks, hate speech does not only create tensions between groups of people, but its impact can affect businesses and create serious conflicts in real life. For this reason, sites such as Facebook, YouTube, and Twitter prohibit the use of malicious expressions as they can endanger the lives of users.

The definition of hate speech is highly controversial and is often used in place of various forms of group-related hostilities such as anti-Semitism, racism, and anti-gypsy [10]. In fact, as [11] pointed out, hate speech is a seemingly complex term, with expressions that can be grouped into such general terms using a variety of derivative and competing concepts. Cyber world hate speech can also change the first intention of social media, which is to communicate and socialize well with other people in the whole world. The "Like" button is made to make people happy, but nowadays, youth can be depressed if they do not get one; it deviates from the original purpose of social media. As a result, the issue that requires this research is to explore the intention of hate speech on Twitter from Malaysians perspectives. Simultaneously, the interaction among human beings should be included as part of this study too.

In an online interaction, people can say what they want without any filter or gatekeeper, either their own posts, tweets, or their comments on some other individual posts. In this way, the whole world can react to each other either in a good or bad way; it is not a problem if good things spread all over the internet, but we are very concerned that bad things, such as hate speeches, will spread, with negative consequences for other people who believe what they see is 100 percent true and real. In sum, throughout this research, the study

wanted to delve into the intention of hate speech on social media, on Twitter. Other than that, this study wanted to explore how hate speech affected other users or victims of hate speech itself.

2. Problem Statement

Watanabe H et al. [12] defined malicious expression or hate speech as the use of offensive, violent, or abusive language; it targets a specific group of people who have something in common, such as gender, ethnicity, race, beliefs or faith, or skin color. Hate speech does not only create tensions between groups of people on the Internet and social networks, but its impact can influence businesses and cause serious conflicts in real life. For this reason, sites such as Facebook, YouTube, and Twitter prohibit the use of malicious language. However, controlling, monitoring, or filtering all content is always difficult. Hate speech and malicious expression is a particular form of offensive language in which the person using it expresses an opinion based on either their biases, personality segregation, racist or radical background, stereotype, or in the worst scenario, just hating the other users.

Oxford Constitutional Law [13] described hate speech as "verbal or non-verbal communication with hostility towards certain social groups, most often racial and ethnicity (racism, exclusion of foreigners, anti-semithism, etc.)" based on gender (sexism), misogyny), sexual orientation (homophobia, transphobia), age (agism), and disability. Therefore, hate speech is considered a global issue in which many countries and organizations are fighting against each other.

Past studies have reported many cases of hate speech all over the world, which can be cancer to the cyberworld society. King and Sutton [14] reported that in the year following 9/11, 481 hate crimes with specific anti-Islamic motivations were committed, 58% of which were committed two weeks after the incident (4% of the risk period). Such evidence shows that unfavorably motivated crimes often occur in close time to electric shock events such as terrorist attacks. As Malaysians and Asians, we are no exception to the pitfalls of hate speech in the cyber world.

Murad D [15] stated that Malaysians living abroad were also victims of racist attacks on Asian communities. There are numerous explanations for this behavior; this is primarily due to misinformation and stigma, alienation, and anti-immigrant sentiments that link the COVID-19 pandemic to China and Asia; these experiences are not limited to the United States. Newcastle University student Sze Qi, 21, was also a target of anti-Asian hatred while studying in the UK. She recalls the incident when she passed a group of teens and one of them coughed fake and said "coronavirus." Last April, a Malaysian student and her Singaporean friend were violently attacked while shopping for groceries in Australia's central business district, Melbourne. The attackers reportedly beat, kicked, and yelled at the "coronavirus" and threatened to kill the two women. In response to this hatred bullying, the #StopAsianHate movement is gaining momentum online by holding physical rallies in the United States, Canada, and Taiwan to raise awareness. UN Secretary-General Antonio Guterres also expressed concern about the increase in violence against Asians during the COVID-19 pandemic.

Meanwhile in Malaysia, some act or policy is needed as transgender people were recently killed in Bukit Tinggi, Klang in 2019, and hate speech to ethnicity, religion and even gender is on the rise [16]. Recent stories at Seafield Temple about land conflicts and their escalation to racial and religious issues show how easily and violently the issues can be transformed into problems of race and religion. In Malaysia, there are no specific laws or policies that can be imposed on those who spread and promote hate speech, especially for the race, ethnic or faith groups, and even for gender.

Article 10 of the Federal Constitution states that all citizens have the right to free speech, but this norm is not absolute. As in other countries, this freedom is subject to some restrictions to protect minorities and people of various ethnic, religious, and sexual orientations, as well as those suffering from other forms of discrimination and aggression. Words, ironically, are sometimes considered a weapon. When archaic incitement is used

against people who disagree or hold opposing views on the government at the time, the law's intended spirit is "to fall into hatred, cause disdain or dissatisfaction with the government." It incites hostility, violence, and discrimination against groups and causes "malicious or interracial hostility." Hate speech, which was created by some problematic groups and has recently been distributed through many communication channels such as Facebook, Twitter, Instagram, and Snapchat, helps to eliminate structural violence, racism, and cyber hatred.

Due to the openness of the Internet and the spread of social media, hate speech has become a prominent topic on the Internet; this will result in experiments with increasing needs and automatic hate speech detection [17,18]; they also stated that in addition to the lack of problems with the traditional terms mentioned above, the lack of a common dataset for conducting research is a difficult obstacle to progress in this area [19].

Based on the past studies, clearly hate speech on social media, especially in Twitter needs some more concern and more studies. Therefore, this study wanted to delve into the intention of hate speech on cyberworld culture based on race, religion, and ethnicity from the Malaysian perspective; this research is relevant to Malaysian social media culture. Azreen H [20] stated that the Malaysian Communications and Multimedia Commission (MCMC) should promote awareness of self-regulation and set a guideline on what constitutes hate speech for the public, as said by a lawmaker, Petaling Jaya (PJ) MP, Maria Chin Abdullah and civil liberties lawyer, Syahredzan Johan, after the institution drop a hotline initiative, where the commission welcomes the public's reports on social media comments that touch on the 3R–Race, Religion and Royal Institution; they also said that education is key because if a person does not have knowledge on what hate speech is, most likely this person would bombard MCMC with various reports just because the content is not up to his liking in the same article. By this research, we can analyze and inform MCMC on the current state of our society nowadays and educate netizens more about hate speech. Based on The Federal Constitution in Article 10, it declares that Malaysians have the right to free expression, but the standard is not absolute. Therefore, this research can make it easier for the government to rebuild or revise an act or article according to hate speech, as it is becoming one of the most harmful things on the Internet after fake news. On the other hand, this research can also contribute a lot to our ministry to conduct an awareness campaign among the netizens to be more polite online. The impactful-ness of the pre-campaign can be measured by this intention of hate speech in the first place.

3. Literature Review

3.1. Hate Speech

"Hate speech," according to [19], is communication or messages that insult an individual or group based on religion, race, caste, sexual orientation, gender, country of origin, or other characteristics. In accordance with [7], malicious language, hate speech or flaming comments is used to identify messages that violate existing legal norms and require government regulation like a publicly shared statement that incites, promotes, or justifies hatred, discrimination, or hostility towards a particular group such as race or ethnicity, beliefs or faith, abilities or disabilities, age, sexual orientation, and gender identity.

Aside from that, Kilvington [8] stated that hate speech is disgusting because it spreads, inflames, and stimulates hatred, violence, and discrimination against individuals or groups due to their protected status; these include "race", ethnicity, religion, gender, sexual orientation, and disability, among other social distinctions. According to the Oxford English dictionary, the term hatred means someone has a strong disgust or dislike, evoking a disliked temperament [2]. Malicious expression or hate speech is a special form of offensive language, where the individual using it is expressing their opinion based on either race, radical background, or stereotype [13].

Thus, by these definitions in previous studies, we can conclude that hate speech is a type of vulgar, obscene, abusive, and abusive language, often expressing hatred and including blasphemous language. Hate messages posted online can harm victims by developing

post-traumatic-stress-disorder such as psychological and pathophysiological symptoms (PTSD)—panic, fear, anxiety, nightmares, threatening intrusive thoughts, contempt [6]. Before this, a lot of researchers focused on how to recognize, detect, and analyze hate speech on social media, especially in the data mining field or how hate speech affected victims in psychology. To differentiate this study and the others, researchers want to investigate more on intentions and motives behind hate speech. The researchers believe that all sentiments and phenomena in previous research are so helpful as they can be proof of the need to do the study.

3.2. Types of Hate Speech

In the untrained eye, hate speech consists of two very common terms: hatred and speech; it refers to emotions and words as well as meanings that can evoke emotions to someone who hates a particular object; it is also usually defined in a broader sense and includes insults, discrimination, dehumanization, demonization, and incitement to violence. According to [21], there are eight types of hate speech as follows: race, behavior, body, class, gender, ethnicity, disability, and religion. The motive of hate speech in [22] study, determine hate speech by Nnamdi Kanu, the leader of the independent people of Biafra; they consider Nnamdi Kanu's speech has five intents or motives which are declaration, commitment, judgment, direction, and assertion. Most of the intent of Nnamdi Kanu's speech is assertive.

In sum, hatred messages or hate speech can be everything or anything that gives negative feedback to other users. Insults or offensive language can be "innocently" negative or sarcasm. In Malaysian cultures, words such as "babi" or "anjing" are so offensive but not "pig" or "dog" in American society. Other than that, calling other individuals using some tribe names such as "batak", "sakai" and "keling" can be offensive and negatively affect other users.

3.3. Intention

Intention is the mental state in which an agent act. One example is the plan to go to the museum tomorrow. An action plan is the content of an intention, and a commitment is an attitude toward that content. Other mental states may be accompanied by the following action plans. As there is no practical obligation to carry out a plan, one can still unintentionally make it. However, only successful intent leads to the intended course of action, whereas unsuccessful intent does not.

Matczak M [23] discovered three types of intention. The three types of intent correspond to the three types of speech: locutionary, illocutionary, and perlocutionary. The initial positional intent is nearly identical to the semantic intent. The intention, for example, is to say a word with a specific meaning and reference. The intent to take specific actions by uttering these words, such as giving orders or making promises, is referred to as spoken intent. Perlocutionary intent is to alter reality by speaking words that influence the behavior of others. The dominant approach in legal theory holds that the desire to legislate is primarily a semantic (linguistic) desire.

3.4. Hate Speech Crime World and Malaysia

Other countries, such as the United States of America [14], reported that in the year following 9/11, 481 hate crimes with anti-Islamic motivations were committed, with 58% of them occurring within two weeks of the incident (4 percent of the risk period). Such evidence indicates that unfavorably motivated crimes frequently occur in close proximity to electric shock events such as terrorist attacks. Other than that, [24] currently in India from 26 May 2022, an inclusive, incredible and colossal India has been in the news for all the worst and wrong reasons in which two apparatchiks of the country's governing party, the Bharatiya Janata Party, or BJP, had made inexcusable incendiary and insulting remarks that have offended more than 10% of the country's minority and the adherents of one of the

world's largest practised religious faiths, Islam, by committing an insult towards Prophet Muhammad (SAW).

According to [25] on 25 March, a video went viral as a Chinese woman falsely accused Malacca Hospital of not treating patients, according to an article in The Center. In addition, a series of online attacks and malicious comments about the woman's ethnicity, such as claiming she left her brain in Wuhan, hoarding, and violations of the Movement Control Order (MCO) have occurred. Between 27 February and 25 March, a total of 706 negative references to Malays/Muslims were registered in the Tweet. "Malas" was used most often, followed by "bodoh". The pinnacle of anti-Malay Muslim mention was on 12 March, when the first few COVID-19 cases were reported by the Tabligh Cluster. Except for anti-Chinese tweets, 506 negative references were recorded during the same period. The highlight of such coverage occurred on March 20, the day after a Chinese cardiologist was discovered jogging in the park against the MCO. The case caused racism on Twitter after his video, discussing with law enforcement officials, became viral. In comparison to other ethnic groups, there were 156 repugnant mentions in anti-Indian tweets during the same time. The highest peak took place on March 22nd, the day a mechanic of Indian descent was charged in court for public intoxication and refusing to cooperate with the police.

There is no specific law in Malaysia that regulates or criminalizes hate speech, especially against ethnicities, religious groups, or skin color. Malaysians have the right to free expression under Article 10 of the Federal Constitution, but this standard is not absolute. As in other countries, some limitations apply to this freedom in terms of protecting minorities, people of different ethnicities, religions, sexual orientations, and others who suffer persecution and aggression.

4. Underpinning Theory

The study is conducted within various theoretical frameworks, one of them is based on the work of [26]. His theory of human engagement and behavior was based solely on an analytical understanding of face-to-face situations in which actors physically coexisted. [27] pioneering work is crucial in understanding how human behavior relates to communication. He was the first scholar to use the dramaturgy metaphor to help him understand himself in social interactions.

Goffman's theoretical model contributed to an understanding of why and how individual and group performance differ in social contexts, particularly when trying to express hate speech. According to [26], individuals acting in a specific social context make "moral demands" on the audience and encourage them to respond in the expected or appropriate manner.

Cultivation is a sociocultural theory of television's part in shaping viewers' notions, beliefs, attitudes, and values [28]. Back then, humans used television as their main tool to get information, news, or public service announcements from the government. Nowadays, when watching football games or concerts, people usually have their mobile phones, tablets, or laptops at arm's length or in the palm of their hands. Often, these devices are used to stream this content. People tend to believe, use ideas and some may follow the things they see on the social media. If hatred messages or hate speech are happening throughout their timeline in social media, it is not impossible if they believe what they watch and see.

According to [8] researchers should try to understand why hatred messages are uploaded online, rather than simply applying Goffman's model to virtual communication. Because the virtual frontstage and virtual backstage are blurred, Goffman's work necessitates extensive rethinking, and redesigning. As a result, researchers combine this model with Gerbner and Gross's cultivation theory from 1976.

In accordance with [29] cultivation is a sociocultural theory that includes three components: media institutions, message production, and message consequences on viewers; however, the effects on viewers are the most popular subject that has been researched. The cultivation effect, in its most basic form, is the correlation between the amount of time people spend watching television and their worldview. Traditionally, the more people who

watch television, the more their worldviews reflect the dominant narrative messages transmitted by television. Consistent with this, the more people who see and read hate speech on social media, the more their worldview reflects the message of the story conveyed by social media. Cultivation theory believes that the potential for storytelling on social media is enormous.

Both models and theories come from different fields as The Dramaturgical Model of Social Interaction comes from psychology and the Cultivation theory is a classic theory from the communication field; it may seem to not be related to each other but in this study, hate speech is in the form of social interaction online and what media has been telling us. In online interaction, people can say what they want without any filter or gatekeeper, either way their own posts or tweets or their comments on some other individual posts. In this way, the whole world can actually react to each other either in a good or bad way; it is not an issue if good things spread all over the Internet, but we are concerned that bad things, such as hate speeches, will have a negative impact on other people who presume what they see is 100 percent true and real; they will always be impacted by what they see on social media, spreading pessimism without questioning whether it is true or not; this could be a viral disease or cancer for the virtual world. Figure 1 shows the relation between The Dramaturgical Model of Social Interaction [26] and the Cultivation theory [28].
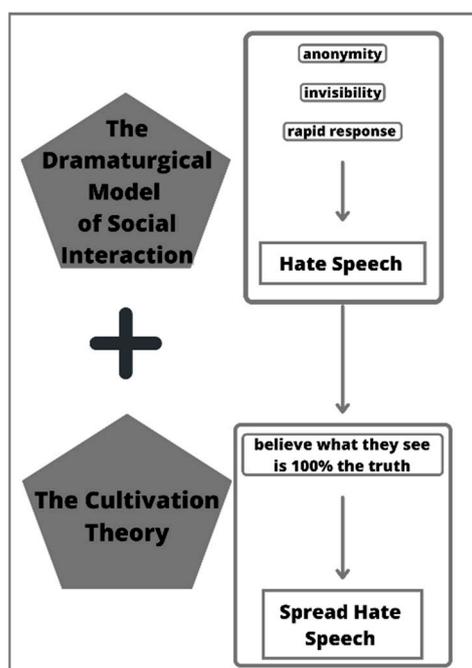

Figure 1. This figure explains how both theories are related to hate speech in the cyberworld.

5. Methodology

The proposed research aims to identify the intentions of hate speech on Twitter from Malaysians perspectives. Through this, the focus of this research is to know the structure of how hate speech affects Malaysians' Twitter lives. The proposed method for this study is to use an interview, which is the qualitative method. Since we want to reach the root of hate speech intent, qualitative research methods are the most appropriate because they are designed to help clarify the behavior and perceptions of the target audience associated with a particular problem [30]. The researcher will interview participants as it is a purely conversational method and encourages participants to receive detailed information [31]. In terms of experiencing using social media, the participants will share how they watch or read hate speech and handle it, and mainly to know the factors that contribute to the occurrence of hate speech.

As we know in the endemic era, we are advised to lessen human-to-human interaction as it can increase the spread of COVID-19. Our target is Malaysians that use social media such as Twitter so all our informants should know how to interact online. We will use video conference platforms such as Zoom Meeting applications or Google Meet applications; it should be easier for everyone, including researchers, to get to interact with everyone without the need to travel around Malaysia.

The Zoom Meeting application will allow you to communicate, chat and discuss things online with other people all around the world. The researchers can connect to those informants that are far from us and it makes it easier in this endemic era; they will also encourage the participants to open their cameras and microphone to make sure that the opinions, as well as facial expressions that the participants want to express, can be recognized in this focus group discussion.

As the target is wide, the kind of informants that should be interviewed or joined the focus group discussion need to be decided. After some meetings with other researchers, it has been decided that the informants, who use Twitter, should come from various backgrounds, such as different races, ethnic groups and academic backgrounds.

This study aims to enhance the validity and reliability of this research paradigm. According to [32], validity and reliability are important because it will help to improve the accuracy of a research work's evaluation and assessment. One of them is to try to make our informants understand why the instruments were used and what our research goals are; this can cause informants to respond to questions in their best minds, increasing the number of ideas that will be used in research findings. Other than that, to make sure our informants give their best opinions, we need to provide some examples or proof of hate speeches that arise on social media, such as Twitter. By this strategy, informants can put their shoes on victims of hate speech as it will increase their empathy.

6. Conclusions

Previous studies paid more attention to how to track and trace hate speech or detect hatred messages in social media. For example, Paschalides D et al. [33] proposed an integrated architecture for detecting online hate speech using multilingual datasets in English, Italian, and German. Corazza M et al. [34] developed a three deep learning method that monitors, detects, and visualizes the occurrence of hate speech using Twitter messages. Essentially, there is little research or article that investigate the motive or intention of hatred message in the cyberworld. In addition, according to recent statistics, Malaysia comprises approximately 20.1 million internet users and 16.8 million social media accounts [3] and cyberbullying was brought to attention when an opinion poll of 28 countries discovered that Malaysia was sixth in the world's cyberbullying rankings, trailing only India among Asian countries [35]; this is very much concerning because it can exacerbate other users and victims of hateful words psychologically and develop post-traumatic stress disorder-like psychological and pathophysiological symptoms (PTSD). Negative consequences include intrusive thoughts of distress, fear, anxiety, nightmares, threats, and injuries [6]. The researchers believe that understanding such extreme negative effects will have both theoretical and practical implications into real-life social interaction.

Malicious expression refers to public statements that express hatred or incite violence towards race, ethnicity, faith, gender or sexual orientation. Online hate diffusion and culture have now become serious problems, prompting several international initiatives aimed at defining the issue and creating effective countermeasures and policies.

Because the goal of this study is to examine the intent of hate speech, researchers believe that human interaction should be studied. A combination of motivation and intention research, the functionalist theory of psychology, the dramaturgical model of social interaction, Erving Goffman [26] and the cultivation theory [28] are used in this research to meet the objective; it is hoped that this study can provide incremental contributions to the hate speech research and provide motives and intentions for other cyberworld users.

References

1. Twitter | History, Description, Elon Musk, and Uses. Available online: https://www.britannica.com/topic/Twitter (accessed on 8 July 2022).
2. Jay, A. 85 Twitter Statistics You Must Know: 2021/2022 Market Share Analysis & Data. Available online: https://financesonline.com/twitter-statistics/ (accessed on 14 January 2022).
3. Kemp, S. Digital 2022: Malaysia. DataReportal—Global Digital Insights. Available online: https://datareportal.com/reports/digital-2022-malaysia (accessed on 15 February 2022).
4. Pattersson, K. The discursive denial of racism by Finnish populist radical right politicians accused of anti-Muslim hate-speech. In *Nostalgia and Hope: Intersections between Politics of Culture, Welfare, and Migration in Europe*; Springer: Cham, Switzerland, 2020; pp. 35–50.
5. Bennett, W.L.; Livingston, S. The disinformation order: Disruptive communication and the decline of democratic institutions. *Eur. J. Commun.* **2018**, *33*, 122–139. [CrossRef]
6. Phoey, L.T.; Chi-Bin, C. Profanity and Hate Speech Detection. *Int. J. Inf. Manag. Sci.* **2020**, *31*, 227–246.
7. Brown, A. What is hate speech? Part 1: The myth of hate. *Law Philos* **2017**, *36*, 419–468. [CrossRef]
8. Kilvington, D. *The Virtual Stages of Hate: Using Goffman's Work to Conceptualize The Motivations For Online Hate*; Leeds Beckett University: Leeds, UK, 2020.
9. Davidson, T.; Warmsley, D.; Macy, M.; Weber, I. Automated Hate Speech Detection and the Problem of Offensive Language. *Proc. Int. AAAI Conf. Web Soc. Media* **2017**, *11*, 512–515.
10. Zick, A.; Krause, D.; Berghan, W.; Küpper, B. Gruppenbezogene Menschenfein-dlichkeit in Deutschland 2002–2016. In *Gespaltene Mitte—Feindselige Zustände: Rechtsextreme Einstellungen in Deutschland 2016*; Dietz: Bonn, Germany, 2016; pp. 33–81.
11. Thurston, N. Back to front truths. Hate library. In *Post-Digital Cultures of the Far Right: Online Actions and of INE Consequences in Europe and the US*; Fielitz, M., Thurston, N., Eds.; The Deutsche Nationalbibliothek: Bielefeld, Germany, 2019; pp. 193–204.
12. Watanabe, H.; Bouazizi, M.; Ohtsuki, T. Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. *IEEE Access* **2018**, *6*, 13825–13835. [CrossRef]
13. Oxford Constitutional Law. Max Planck Ensyclopedia of Comparative Constitutional Law. Available online: https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e130?prd=MPECCOL (accessed on 20 June 2022).
14. King, R.D.; Sutton, G.M. High Times for Hate Crime: Explaining the Temporal Clustering of Hate Motivated Offending. *Criminology* **2013**, *51*, 871–894. [CrossRef]
15. Murad, D. Caught in the Hate-Demic. *The Star*, 18 April 2021.
16. Lih, Y.B. Death of Transgender Woman in Malaysia Sparks Fears of Rising Hate Crime. U.S. Available online: https://www.reuters.com/article/malaysia-lgbt-crime-idUSL8N1YM12C (accessed on 20 June 2022).
17. Kiela, D.; Firooz, H.; Mohan, A.; Goswami, V.; Singh, A.; Ringshia, P.; Testuggine, D. The hateful memes challenge: Detecting hate speech in multimodal memes. *Adv. Neural Inf. Processing Syst.* **2020**, *33*, 2611–2624.
18. Malmasi, S.; Zampieri, M. Detecting hate speech in social media. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Varna, Bulgaria, 2–8 September 2017; pp. 467–472.
19. Schmidt, A.; Wiegand, M. A survey on hate speech detection using natural language processing. In Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Valencia, Spain, 3 April 2017; Association for Computational Linguistics: Valencia, Spain, 2017; pp. 1–10.

20. Azreen, H. MCMC Urged to Promote Awareness On Self-Regulation, Hate Speech Online. The Malaysian Reverse. Available online: https://themalaysianreserve.com/2019/08/19/mcmc-urged-to-promote-awareness-on-self-regulation-hate-speech-online/ (accessed on 20 June 2022).
21. Mondal, M.; Silva, L.A.; Benevenuto, F. A measurement study of hate speech in social media. In Proceedings of the 28th ACM Conference on Hypertext and Social Media, Prague, Czech Republic, 4–7 July 2017; pp. 85–94.
22. Alabi, T.O.; Ayeloja, A.K. Hate speech and security challenges: A pragmatic study of Nnamdi Kanu's speeches in the south-eastern Nigeria. *Int. J. Engl. Res.* **2019**, *5*, 1–9.
23. Matczak, M. Three Kinds of Intention in Law making. *Law Philos.* **2017**, *36*, 651–674. [CrossRef]
24. Santhananaban, M. Hate speech has no place anywhere. *The Sundaily*, 10 June 2022.
25. Jia, V.T.; Omar, N. Like a Virus: How Racial Hate Speech Looks Like in Malaysia During the COVID-19 Pandemic. *The Centre*, 2 April 2020.
26. Goffman, E. *The Presentation of Self in Everyday Life*; Penguin Books: London, UK, 1959.
27. Goffman, E. *Frame Analysis: An Essay on the Organization of Experience*; Harvard University Press: Harvard, MA, USA, 1974.
28. Gerbner, G.; Gross, L. Living with television; The violence profile. *J. Commun.* **1976**, *26*, 182–190. [CrossRef] [PubMed]
29. Shrum, L.J. Cultivation theory: Effects and underlying processes. In *The International Encyclopedia of Media Effects*; Roessle, P., Ed.; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2017.
30. Creswell, J.W. *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2013.
31. Mathers, N.; Fox, N.; Hunn, A. *Using Interviews in a Research Project*; Trent Focus Group: Sheffield, UK, 2002.
32. Mohajan, H.K. Two criteria for good measurements in research: Validity and reliability. *Ann. Spiru Haret Univ. Econ. Ser.* **2017**, *17*, 59–82. [CrossRef]
33. Paschalides, D.; Stephanidis, D.; Andreou, A.; Orphanou, K.; Pallis, G.; Dikaiakos, M.D.; Markatos, E. Mandola: A big-data processing and visualization platform for monitoring and detecting online hate speech. *ACM Trans. Internet Technol. (TOIT)* **2020**, *20*, 1–21. [CrossRef]
34. Corazza, M.; Menini, S.; Cabrio, E.; Tonelli, S.; Villata, S. A multilingual evaluation for online hate speech detection. *ACM Trans. Internet Technol. (TOIT)* **2020**, *20*, 1–22. [CrossRef]
35. Rosli, J. Malaysia Sixth-Worst in Global Cyber-Bullying Ladder, Survey Shows. Available online: https://www.malaymail.com/news/malaysia/2018/10/27/malaysia-sixth-worst-inglobal-cyber-bullying-ladder/1687181 (accessed on 15 June 2022).