Check for updates

RESEARCH ARTICLE

REVISED **The current situation of online harassment in Japan and its countermeasures:**

The E x a m p l e o f "Flames" of Tweets by Newly-employed Families on Twitter[version 2; peer review: 2 approved]

REVISED **Online harassment in Japan: Dissecting the targeting of a female journalist [version 2; peer review: 2 approved].**

Aki Nitonami (iD) 1, Mitsuo Yoshida (iD) 1, Yukie Sano 2 (iD)

1University of Tsukuba, Department of Business Science, Japan
2University of Tsukuba, Japan

Aki Tonami (iD) 1, Mitsuo Yoshida (iD) 1, Yukie Sano 2 (iD)

1Faculty of Business Sciences, University of Tsukuba, 3-29-1 Otsuka, Bunkyo, Tokyo, 112-0012, Japan
2Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan

Open Peer Review

Approval Status ✔ ✔

	1	2
version 2 (revision) 03 Feb 2022		✔ view
version 1 17 Nov 2021	✔ view	? view

summary
Social media "flames" are no longer an exaggeration to say that they are a
However, there are few studies on flaming in the Japanese language
environment based on the viewpoints of the victims. This paper
introduces the concept of "online harassment" and outlines the current
situation based on a review of previous studies in Japan and abroad.
Next, we analyzed a case of "inflammation" of tweets by a female
journalist on Twitter, and observed that there were three groups of
users who conducted harassment: an influencer group, a group of users
who participated in inflammation in response to the influencer's
whistle, and a group of users who trolled, and each group harassed a
human user in a different way. The first group of users was the flame-
throwers who responded to the flames, and the second group was the
trolls. Finally, we will discuss measures that should be taken by
individuals and organizations.

1. **Yasuko Kawahata**, Rikkyo University, Tokyo, Japan

2. **Shinichi Yamaguchi** (iD), International University of Japan, Niigata, Japan

Any reports and responses or comments on the article can be found at the end of the article.

Abstract
Harassment on the Internet, particularly on social media such as
Twitter, has reached a level where it can, without exaggeration, be

characterised as a real-world societal problem in Japan. However, studies on this phenomenon in the Japanese language environment, especially adopting a victim-centric In this paper, we incorporated the concept of online harassment and reviewed existing studies about online harassment from Japan and abroad. We then conducted a detailed case analysis of the "flaming" of a female journalist and those who targeted her on Twitter. Based on our analysis, we observed that there were three layers of users who targeted the journalist: influencers, users who responded to the

Each harassed the journalist, but in a different manner. Given Japan's particular difficulty of imposing domestic regulations on social media companies that are mostly from abroad, we propose and describe possible measures that individuals and their employers should consider taking. Each harassed the journalist, but in a different manner.

Keywords.

social media, online harassment, Twitter, policy, social media, online harassment, policy, COVID-19

This article is included in the Japan Institutional Gateway gateway.

Corresponding author: Aki Tonami (tonami.aki.ka@u.tsukuba.ac.jp)

Author roles: Tonami A: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing; **Yoshida M**: Formal Analysis, Investigation, Methodology, Writing - Review & Editing; **Sano Y**: Data Curation, Investigation, Visualization, Writing Visualization, Writing - Review & Editing

Introduction.

In recent years, advances in digital technology and the proliferation of social media[1] (SNS) has made it possible for many people to easily transmit and gather information. The frequency with which people use the Internet has also increased markedly in response to the spread of the new coronary infection (COVID-19) outbreak in 2019 and the government's declaration of a state of emergency, as people refrain from going out and meeting face-to-face as much as possible (Morishita, 2021) At the same time, there have been a number of incidents related to problems on the Internet, especially communication failures on social networking services (SNS). It is no exaggeration to say that so-called "online flames" have become a social problem. Flames" refers to the phenomenon of a flood of critical comments on social media about what a person has said or done (Yamaguchi, 2015)

However, most of the existing studies on flame wars deal with the phenomenon of information on social networking services (SNS), and studies based on the perspectives of parties affected by flame wars are rare, especially for SNS in the Japanese language environment. Therefore, this paper introduces the concept of online harassment when examining the phenomenon of flame wars, and asks the descriptive question, "What is the current state of online harassment in the Japanese language environment? Online harassment is defined as "repeated inflammatory comments or hate speech on the Internet ("trolling" and "trolling") cyberstalking, physical intimidation, publishing sexually explicit images without consent ("revenge pornography") disclosing personal information ("exposing " and "doxing")" (PEN America, n.d.-b).

In the following, we first provide an overview of the current state of online harassment based on a review of previous studies, followed by a case analysis of "flames" of tweets by female journalists and online harassment that occurred on Twitter, a social networking service used by approximately 55 million people in Japan (Statista Research Department, 2021). Finally, we will discuss the measures that individuals and organizations should take to prevent online harassment. Please note that this is a qualitative study and is not intended to be empirical or hypothesis-testing.

Current State of Online Harassment

At the time of writing this paper, existing research on online harassment is rare. A search of the academic database Web of Science using "online harassment" as a search term across all text fields yielded 1,360 articles. A pediatrics article on "Internet harassment" of young people, published in 2006, may have been among the first scholarly articles dealing with online harassment (Ybarra et al., 2006). On the other hand, this number is much smaller than, for example, the 30,225 results of a search for the term "Twitter".

Online harassment is a negative impact that digital technologies and social networking sites have had on society, but there are not many even large-scale studies dedicated to this behavior, perhaps in part because existing studies on digital technologies have been optimistic about these negative changes (Fung et al., 2013; Jankowicz et al., 2021). In addition, since the problematic phenomenon of online harassment began to be rapidly recognized, mainly in the English-speaking world, triggered by the 2016 U.S. presidential election, as discussed below, the discussion of survey analysis also tends to be policy-oriented, with documents containing policy recommendations made by public organizations or co-authored with NGOs, etc. These studies are disinformative. These studies are also often disinformation[2] These studies are often conducted in the areas of disinformation, online abuse, and human rights (Jankowicz et al., 2021). For example, from a human rights perspective, the "Report on the promotion of gender justice and freedom of opinion and expression" published in July 2021 by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression addresses harassment, especially against women, in the Internet environment (Khan, 2021). (Khan, 2021). In Japan, online harassment was initially seen as a transfer of bullying among children to the Internet environment (Sho et al., 2018) Shin and Hamada (2021) analyzed Twitter replies to three female local council members and noted that there were two types of replies: those that denied or ridiculed the council members' statements or political activities, and those that tried to attract their attention by (persistently) expressing instructions or support for the council members.

So, what kind of harassment is taking place in the on- online space, and to what extent?

In the following, we attempt to summarize the results of a few research studies.

[1] Individuals, communities, and organizations create, co-create, modify, share, and engage with easily accessible user-generated content
and web-based services to enable collaboration, connection, interaction, and community building (McCay- Peet & Quan-Haase, 2017).

[2] information-manipulation cyberattacks (Nagasako, 2021).

According to a 2020 survey by Plan International, an international NGO that promotes children's rights, 25 % of 501 young women aged 15-24 in Japan said they had experienced some form of online harassment on social networking sites (Plan International, 2020 2020). On the other hand, a comparison of survey results from 31 countries revealed different understandings of the concept of online harassment in different countries, and that it was not understood at all by some respondents.

According to a 2021 survey report by the U.S.-Pew Research Center, one in four Americans have experienced online harassment, and of these, half said they were harassed because of their political opinions (Vogels, 2021). In addition, compared to similar surveys in 2014 and 2017, the number of respondents who said they had been subjected to more serious online harassment, including physical threats, stalking, ongoing harassment, and sexual harassment, increased.

Regarding the Japanese language environment, in a survey of 770 men and women in their 20s to 60s who use SNS nationwide by Internet provider BIGLOBE, Inc.[3] In the survey, 4.5% answered "often" and 13% answered "sometimes" when asked if they had been slandered by others on SNS (BIGLOBE, Inc., 2020). Among these, those in their 20s had the highest percentage of being slandered by age group, with 10% saying "often" and 18.9% saying "sometimes. It is important to note, however, that slander here is not the same as the definition of online harassment. The term "slander" is a combination of the two words slander (to speak ill of another person) and defamation (to speak ill of another person with no basis in fact and to damage his/her reputation) (Sanseido, 2013). Flaming, as mentioned above, is a phenomenon in which critical comments, including slander, flood social media, while online harassment is an online act of harassment, including slander, which is sometimes concentrated and can become a flaming phenomenon.[4]

Apart from attempts to uncover the full extent of online harassment, there are findings from studies conducted with an approach that prioritizes helping victims recover. PEN America, the U.S. chapter of the International PEN Club, an organization of writing professionals, took a serious look at the serious harassment that reporters and writers are experiencing on social networking sites, and launched a research project called the Online Harassment Field Manual (PEN America, n.d.-c). The organization conducted a survey in 2017, which received responses from 230 people who had been subjected to online harassment (196 of whom indicated their gender, 136 were female, 52 were male, and 8 were other) indicating that online harassment occurring on the Internet has a serious impact on the real lives of the victims It was clear that online harassment occurring on the Internet has had a serious impact on the real lives of the victims. For example, 67% of the respondents "reacted seriously by feeling that they or someone close to them was in danger, refraining from publishing their work, or permanently deleting their SNS accounts"; 64.3% "suspended their SNS because of the harassment"; 62.3% "reported that online harassment had affected their personal life, physical, psychological physical, or mental health" (PEN America, n.d.-b). As for the reasons for the harassment, 53.5% said they were harassed "for expressing their political views" or "for expressing personal opinions," 38.9% said they were "targeted because of their gender or gender identity," and 31% said "appearance (lookism)," indicating that online harassment is a result of individuals expressing their opinions on social networking sites This indicates that online harassment is being perpetrated either because of the opinions expressed by individuals on social networking sites, or because of the gender or identity of the speaker.

In addition, even in academic circles, where the use of social media is often encouraged for research publication activities, to expand networking among researchers, and to help facilitate communication with students, researchers are frequently the victims of online harassment, with detrimental effects Gosse et al. Gosse et al. (2021) noted that although online harassment is viewed as a job, identity, or scholarly requirement, it is complicated by the entanglement of gender and appearance (lookism) factors.

According to Sobieraj (2020), anyone can be a victim of online harassment, but it is especially aggravated There are certain attributes that make them easy targets for harm. Namely,

1) Minority,

2) Feminist (not bound by old gender values)

3) Opinions on politics, sports, diplomacy, defense, cybersecurity, and other male-dominated areas

---

[3] The survey found that "people who defame others by spreading baseless bad words, or who hurt others with words that denigrate their personalities, are the most likely to be involved in the following behaviors.
  and is defined as "and.

[4] Other similar terms include cyberbullying and cyberbringing, but at the time of writing, definitions of each term were not available.
  It appears that this has not yet been determined. Further discussion and organization are desirable.
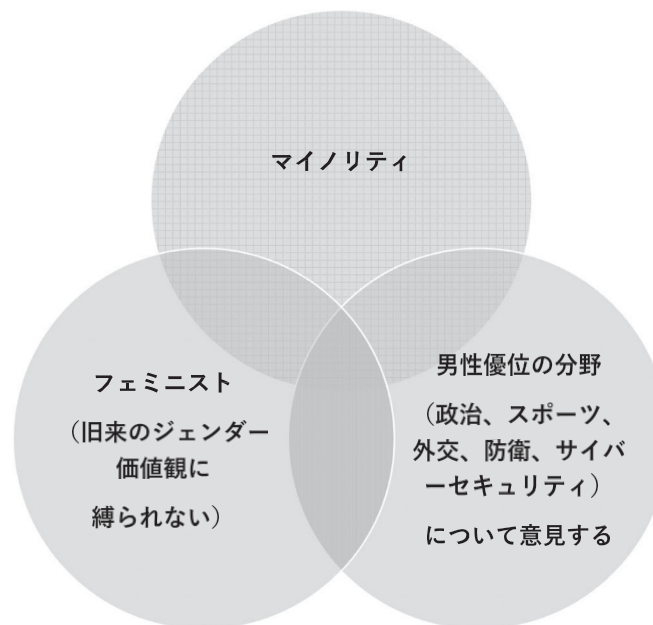
Figure 1: Attributes Susceptible to Online Harassment (from Jankowicz *et al* (2020)).

are more likely to be targeted for harassment, and are most likely to experience harassment in the common areas of these attributes (Figure 1) Note that in the European survey, 58.2% of female legislators responded that they had experienced online harassment on social networking sites, and that aspects of being a minority in parliamentary politics, particularly being younger (under 40) and then being a member of an opposition party, increased the risk of being harassed (Inter-Parliamentary Union, 2018; Shen and Hamada, 2021)

According to a report by Online SOS, a support group for victims of online harassment in the United States, online harassment has existed since the 1990s, when the Internet was commercialized, but the 2014~2016 "Gamergate"[5] period, a series of harassment trends were established in the 2016 U.S. presidential election (Lee, 2019). In other words, there is a mechanism by which malicious actors harm their targets (victims)through the use of a variety of measures, through numerous media, and in multiple locations (Figure 2). Thus, recent studies have highlighted the problem that even acts of harassment in the online space can have a profound impact on victims and that harassment and various acts of violence are not connected to the online space but to the offline, or real world (Berry, 2019)[6]

How are SNS platforms dealing with such online harassment? Jankovic et al. (2021) found that SNS platforms with a large number of users, such as Facebook, Twitter, and YouTube, have incorporated anti-harassment guidelines into their terms of use and "community guidelines," but they are uniform and insufficient compared to the reality of online harassment. However, they are uniform and inadequate compared to the reality of online harassment. In the aforementioned PEN America survey, 53.1% of the 230 respondents who had experienced online harassment reported the harassment to the platform, and 70.8% of them said that the platform's response was not helpful at all.

These existing studies analyzed online harassment mainly in Western language environments, and they are cursory because of the mixed definitions of harassment and of who perpetrated and was victimized by harassment.

---

[5] An incident that began in Europe and the United States around the summer of 2014, in which a conflict between gaming enthusiasts who advocate freedom of expression and feminists who call for restrictions on expression from the perspective of discrimination against women intensified on the Internet. Extreme online harassment, including leaking of personal information, death threats, and intimidation, took place against certain female programmers.

[6] Figure 2 points out that online harassment is taking place across multiple platforms other than Twitter. The analysis of such actions is beyond the scope of this paper's analysis and should be the subject of future research.

有害なアクターが	様々な方策を 使用し	多くの媒体で	複数の場所で	ターゲットに 危害を与える
• 匿名人物 • 同僚 • 元同僚 • 元友人 • 元パートナー • パートナー • 別離した家族 • ファン • 友人や家族 • 集団 • イデオローグ • 雇われトロール （荒らし） • ニセ匿名人物 • 読者 • ストーカー	• いじめ • 調整されたターゲティング • ドキシング（晒し） • 不愉快なコメント • なりすまし • 扇情的なコメント • 集団ハラスメント • 同意のないポルノ画像 • ストーキング • 継続的なハラスメント • 警察への通報 • 暴力行為の脅迫 • トローリング • 意図的に恥ずかしい思いをさせる • 動揺させるコメント	• ブログ記事 • コメント • ダイレクトメッセージ • Eメール • 画像 • ミーム • ソーシャルメディアのポスト • テキストメッセージ • モバイルペイメントサービス • ビデオ	**ソーシャルメディア** • Facebook • Twitter • Instagram • Snapchat **掲示板** • Reddit • 5ちゃん • 4, 8chan • WeChat • Yahoo!コメント **個人メディア** • Eメール • LINE • テキストメッセージ • ダイレクトメッセージ **一般向けメディア** • コメント部分 • YouTube • Google • Wikipedia/pedia **ニッチなプラットフォーム** • フリマアプリ • 出会い系アプリ	• 不安 • 降伏 • 評判の悪化 • 死 • 不和 • 排除 • 恐怖 • 孤独 • 身体的暴力 • 口封じ • トラウマ

Figure 2. based on Components of Online Harassment (Lee, 2019).

It is undeniable that the "online harassment" is a problem in the Japanese language environment. Therefore, in this paper, through case analysis, we will examine the online harassment in the Japanese language environment. Attempt to detail an example of the damage and mechanism.

Case Analysis : "Flame" incident involving a reporter's tweet from Company X

In this paper, a qualitative approach was used to analyze the case studies. The research paradigm is based on interpretivism. The Twitter account analyzed is the account of a real-name female reporter (hereinafter referred to as "Ms. Z") who belongs to a major media company (hereinafter referred to as "Company X"), has approximately 20,000 followers, and has been certified by the Twitter company as a prominent person. The author uses Twitter and has been following the account for some time, and thus had confirmed that the account had previously been the victim of a "flame war. The case under analysis was a firestorm triggered by a tweet regarding the number of deaths caused by the new coronavirus, which was made at a time when Japan had recorded the highest number of new infections in its history (Nihon Keizai Shimbun, 2021a) This may have had something to do with the fact that the tweet was retweeted by many users, including quotes, as described below, and was seen by the authors, so it was selected as the case study for analysis.

At 7:12 p.m. on August 3, 2021, Mr. Z tweeted that he was saddened when he saw tweets that described the number of deaths caused by the spread of the new coronary disease as small, and that regardless of the number, each person who died had a life, and that it was not right for celebrities in particular to trivialize it by calling it "small. It's not right to trivialize it, especially by celebrities.

After this tweet was posted, critical tweets were concentrated for about three days, resulting in a state of inflammation. In this paper, we analyzed 2,817 direct replies (1,175 direct replies, 1,499 retweets of these replies, and 143 quoted retweets) and 5,595 quoted retweets of this tweet (hereafter quoted RTs) that reached Mr. Z's account directly between August 3 and 7, 2021. The tweets for data analysis were analyzed using the ExportData.io service (to:[Mr. Z's account ID]) until: 2021-08-07 since: 2021-08-03 and Query: https://twitter.com/[Mr. Z's account ID]/ status/ 1422500544032038919, Result type: mixed, Language: en,

retrieved on August 7, 2021. Free alternatives with equivalent functionality include Twitter's official API (https://developer.twitter.com/en/products/twitter-api, or https://developer.twitter.com/en/docs/ twitter-api/tools-and-libraries).
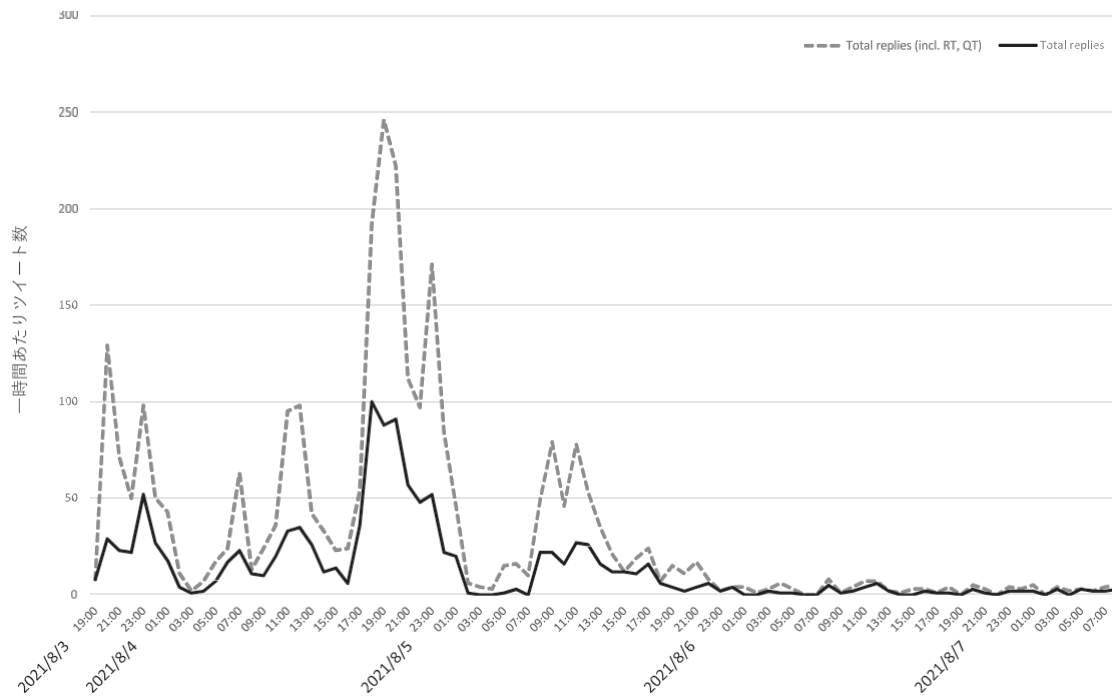
Figure 3. number of direct replies and their retweets and cited RTs per hour.

The reason for focusing on direct replies is that even if a quoted RT is spread, the quoted user is not notified and can observe the situation quietly, but if a direct reply or quoted RT is made to the quoted user, he or she is notified, and especially if the comment is critical of the quoted user's identity or something other than the content of the original tweet, he or she is confronted with harassment. However, if a direct reply or quoted RT is made to the quoted user, the quoted user will be notified and can observe the situation quietly, but if the comment is critical of something other than the content of the original tweet, such as the identity of the quoted user, the quoted user will face harassment, which can cause significant emotional distress.

Figure 3 graphs the number of direct replies to Mr. Z's account and the number of tweets per hour that retweeted or quoted RTs of direct replies. It can be seen that the number of direct replies per hour peaked at 11:00 p.m. on August 3, declined, and then surged again from 4:00 p.m. on August 4, continuing to receive close to 100 replies per hour until 9:00 p.m. on the following day. Using the qualitative analysis software NVivo Windows (version released in March 2020) we automatically classified the sentiments of the direct replies, as shown in Figure 4, and found that the majority of the direct replies were critical in nature. The KH Coder 3 and ML-Ask, a Python library, are free alternatives that perform the same function. These can be used to analyze the frequency of words appearing in tweets and their sentiment.

Further manual classification of the tweets revealed that of the 1,175 direct replies, 147 were positive and 1,022 were critical. A word cloud was also created using the same software, as shown in Figure 5. The figure shows that those who sent direct replies asserted that "the number of deaths from (novel) coronavirus infection in Japan is small."

The direct replies were manually classified as[7] Table 1. The most frequent claim to the content of the original tweet was that "it is fair to say that the number of deaths is low" (216 replies) followed by "there are other important issues (new coronary infections)" (170 replies) These are not formal errors.

---

[7] The same tweet may contain multiple codes. In this paper, we use "generative coding," in which codes are attached freely in an exploratory manner while reading the data.

The "coding generative coding" was performed. In this coding, which is often employed in qualitative research,

interpretation plays an important function and is considered proactive. For this reason, it is rare to have multiple coding workers, and differs from "standardized coding," in which codes are selected and attached from a selected set of codes and multiple coding workers are assigned to evaluate the degree of agreement among them (Otani 2019)
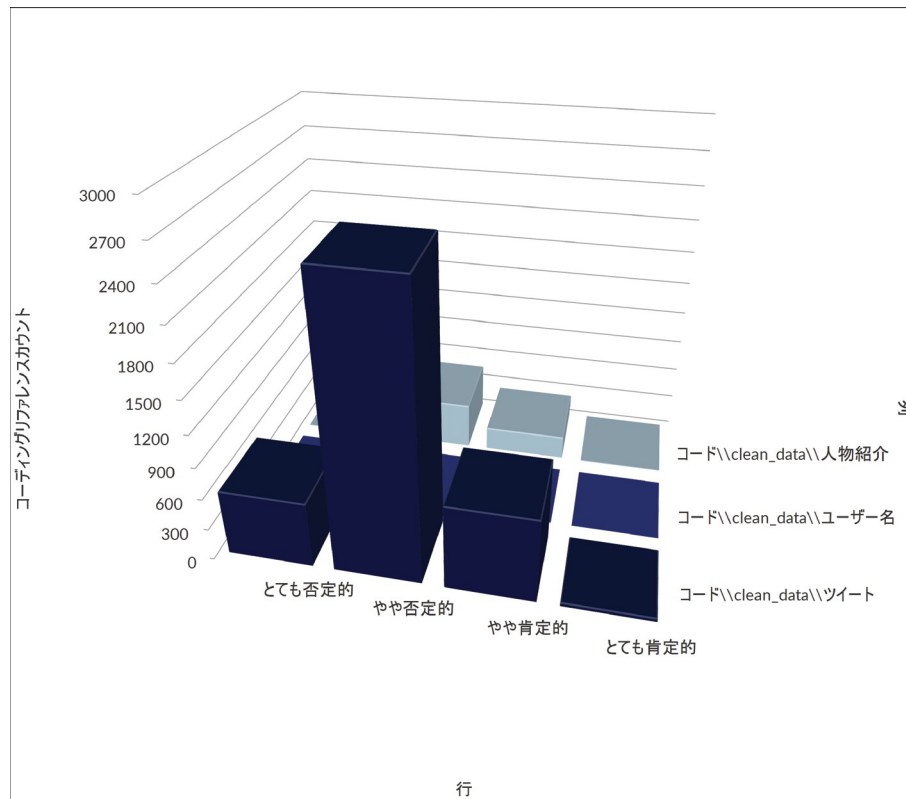
Figure 4: Results of automatic emotion coding.

Figure 5. word cloud (proper nouns are anonymized).

Table 1: Classification and number of direct replies.

There are other important matters ti		170
	Only COVID-19 should not be singled out.	70
	deaths due to economic deprivation	49
	HPV vaccine	10
	Israel	1
It is fair to say that less is more.		216
It's inflammatory.		71
The numbers should be "properly" reported.		30
misogynistic		150
	Don't get emotional.	124
	Poem."	16
criticism of the administration		24
defending the administration		28
Company X		186
criticism of the media		177
it can't be helped		39
personal attack or denunciation		195

The comments are based on "whataboutism," a defense of hypocrisy and hypocrisy. In addition, 71
The direct replies were critical of the original tweets, which they viewed as inflammatory.

Many of the direct replies criticized or slandered Mr. Z's attributes rather than the content of the original tweets. For example, there were 195 personal attacks that had nothing to do with the content of the original tweets, such as "Shame on you," "I despise you from the bottom of my heart," "Shut up, scum," and "How can you be a reporter if you are like that?" and 186 criticism of Company X,
There were 177 criticisms of the media and 150 comments that were misogynistic, such as "Don't be emotional. The "women are emotional" is a kind of gender stereotype (Morinaga, 2013) a simple and fixed idea about the personalities, abilities, roles, and behaviors of men and women, such as "women are kind," "men are strong," "women are caring," and "men can do numbers" (Weigard et al., 0 These prejudices distort perceptions of others and lead to sexism, especially benevolent sexism (Glick & Fiske, 1996, 1997, 1999, Correll et al., 2020). Even in the online environment of Japanese, it has been noted that speakers' opinions are suppressed by attacking their speech performance as "emotional," especially toward women (Tanaka, 2020) For these reasons, we determined that "emotional" was not a personal attack, but a misogynistic, i.e., discriminatory remark against the group to which Ms. Z belonged (or was perceived to belong).

Thus, the reasons for Mr. Z's online harassment were that he expressed his political views, that he expressed his personal opinions, that he is a woman, and appearance (lookism), which were also found in the PEN America survey results: individuals expressed their opinions on social networking sites, or that the speaker's It can be seen that this is done on the grounds of gender and identity. In addition to this, we found that an aspect of Mr. Z's being a reporter for Company X was a reason for identity attacks unique to the Japanese Internet space.

Next, we visualized and analyzed the hourly volume of direct replies to Mr. Z and RTs quoting the original tweet, and inferred that users who engaged in harassment through flame wars were divided into three groups. The three groups of users who engaged in harassment through flame wars were (1) influencers (users who were involved in high frequency flame wars) (2) users who participated in flame wars in response to influencers' "dog-whistles," and (3) users who trolled.

A detailed analysis of 77 users who were involved in more than 5 flameouts revealed that there was a close follow/follow relationship between these users, with users who were involved in more than 5

flameouts having a large number of followers and tweeting about the same flameout topic, The analysis of 77 users involved in 5 or more flame wars showed that there was a tight follow/follow relationship between these users, and a resonant structure of information, with users involved in high frequency flame wars having a large number of followers and tweeting about the same flame topics.
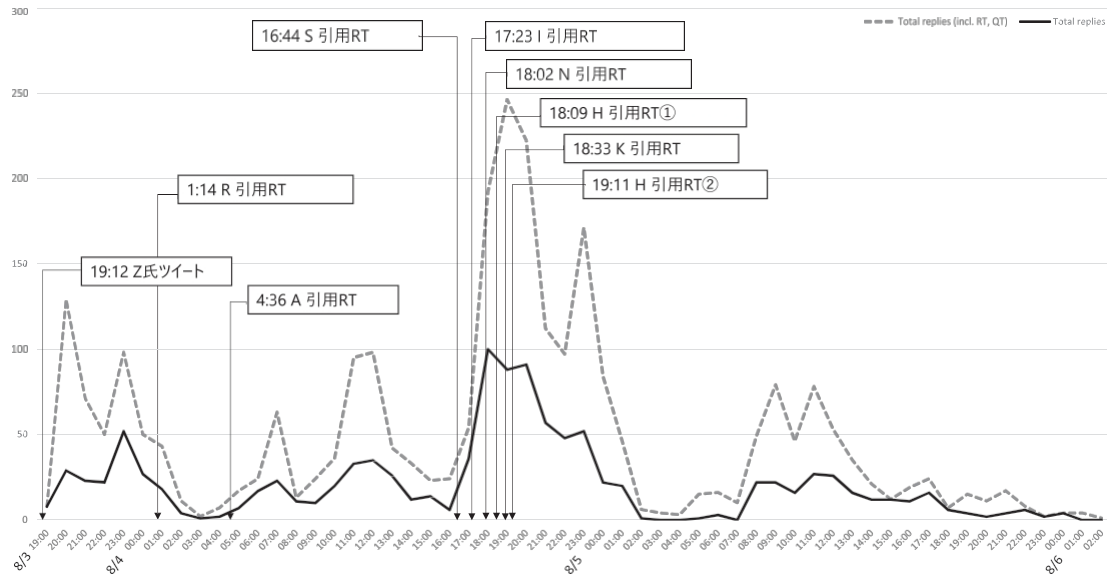
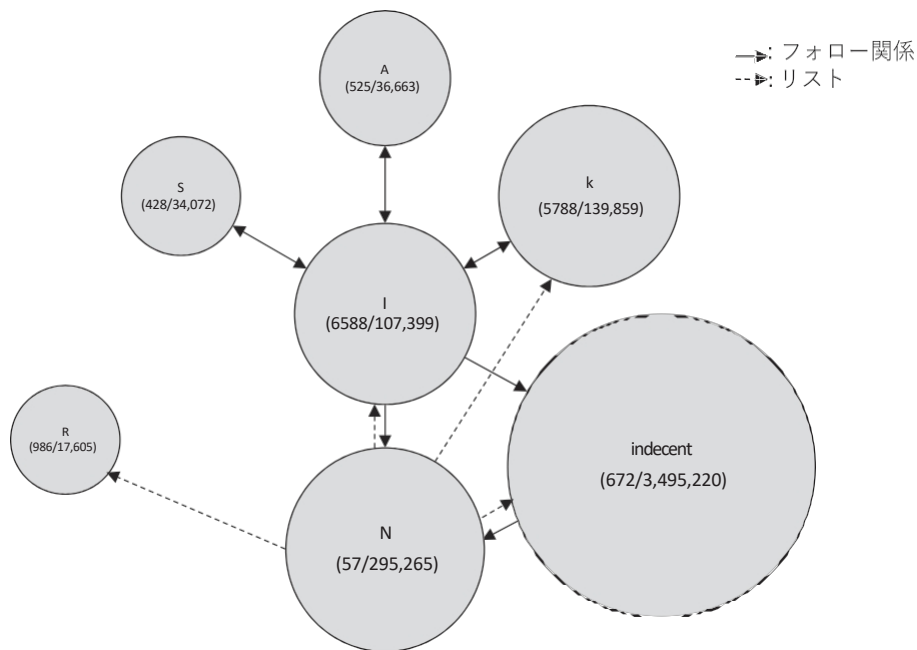Figure 6. direct replies and influencer citations per hour RT.

Figure 7. Influencer follow/follower relationships (number of followers in parentheses).

Inthe case of Mr. Z's flaming incident, users with a large number of followers and specific political tendencies formed a follow/follow relationship and sequentially engaged in quote RTs of Mr. Z's tweets among users (Figure 6 and Figure 7). Although it is difficult to consider these users as a group of users involved in high frequency flame wars, it can be inferred that the quoted RTs of users with a large number of followers spread widely across Twitter through their followers (Fig. 8).

(2) A group of users who respond to the "dog whistle" of the influencers in (2), but a dog whistle is "an aggravating or harmful meaning.

Specific markers to perpetrator groups on the Internet, using double or coded words or symbols with
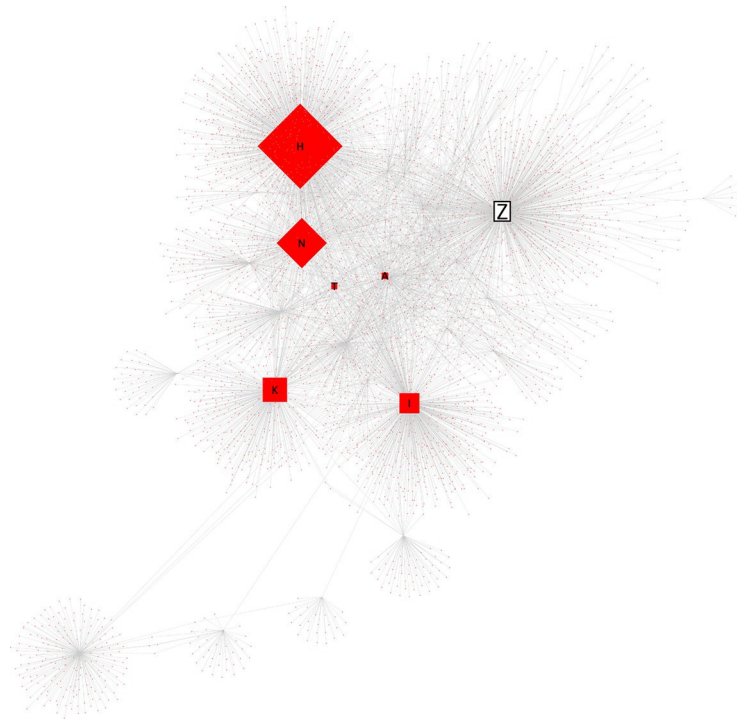
Figure 8: Citation RT diffusion network diagram by influencers.

It means "the act of signaling a target to attack the target" (PEN America, n.d.-a). Of the aforementioned influencers, Mr. H, who has approximately 3.5 million users among his followers, quoted RTs of Mr. Z's tweets on two separate occasions. The first RT was made at 6:09 p.m. on August 4, 2021, and said that the reporter was doing a lousy business of earning circulation by raising readers' fears that the death toll from the new coronavirus was high, and that he wished he would study more science and write articles calmly and logically, but since it would be impossible, the media concerned itself should cease to exist. I wish they would learn more about science and write calmly and logically, but since that would be impossible, I wish the media itself would cease to exist. The second quoted RT was made at 7:11 p.m. on the same day, stating that Mr. Z's tweet was justified in making money by raising fears about the new coronavirus and was the worst.

The common claim of these two tweets is "don't incite". As mentioned earlier, 71 of the direct replies criticized the original tweet for being inflammatory, but a chronological check of the time period in which these occurred shows that although there had been only a total of 6 direct replies with this content before, there were 7 at 6 p.m. when Mr. H made the first quoted RT, 15 at 7 p.m. when he made the second quoted RT, and 11 at 8 p.m. (Figure 9). The number of direct replies increased sharply to 7 pm when Mr. H sent his first quoted RT, 15 at 7 pm, and 11 at 8 pm (Fig. 9) There were also three tweets stating that they came from Mr. H's tweets, indicating that they went out of their way to directly and intentionally attack Mr. Z's account in response to the influencer's dog whistle. According to an existing study of the flame phenomenon by Yamaguchi (2015, 2016, 2020), most people give a sense of justice as the reason for joining the flames, and it is assumed that this group of users is the majority of those who joined the flames.

From the analysis of direct replies, apart from influencers and users who were presumed to have responded directly or indirectly to influencers' calls for help, we found a group of users who (iii) trolled (repeatedly posted inflammatory comments or hate speech). First, out of 195 tweets of personal attacks, 12 users repeatedly sent replies to Mr. Z with personal attacks more than twice. Figures 10 and 11 graph the number of direct replies with critical content and personal attack replies per hour, but the quantities of the two do not necessarily correspond to each other. The time period with particularly high numbers of personal attack comments per hour (11:00 a.m., August 4, 2021)
～When compared to the number of critical tweets per hour in response to the earlier influencer's

dog-whistle (August 5, 2:00 a.m.), there is no particular similarity in the trend of increase or decrease (Figure 12). What can be inferred from these results is the possibility that there is a group of users who regularly monitor Mr. Z's comments and troll him when the time comes. In fact, a self-proclaimed media watcher with more than 7,000 followers tweeted on March 18, 2021, that there are some low-quality and ideological reporters at Company X, and listed them as a list of "low-quality" reporters at Company X.
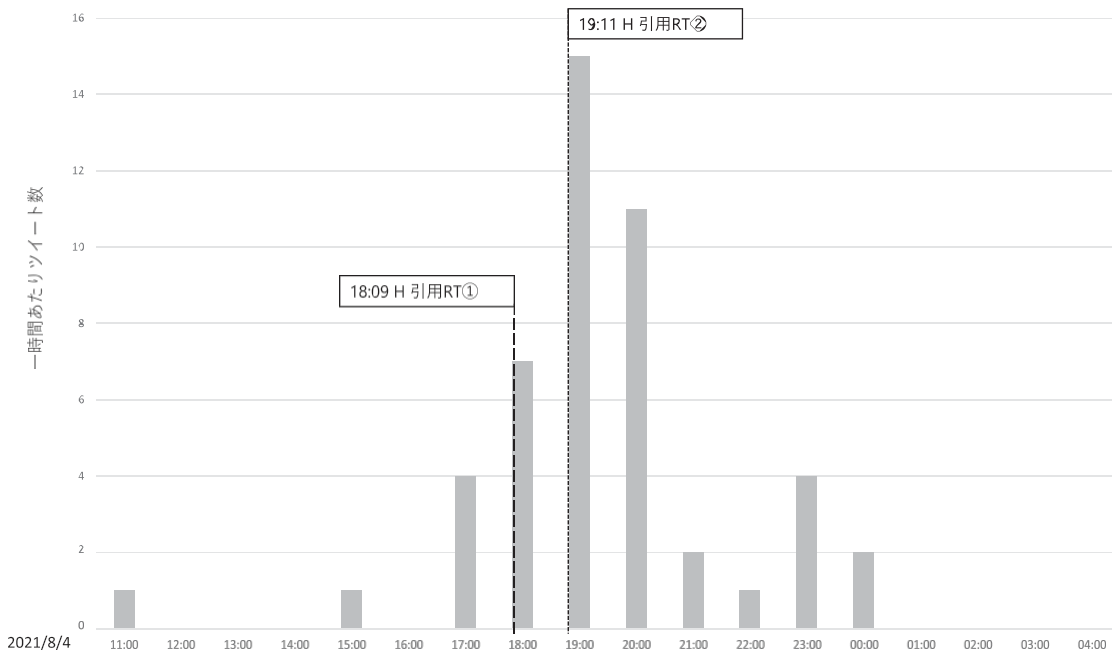
Figure 9. hourly direct replies and influencer quoted retweet time for "don't agitate" content .
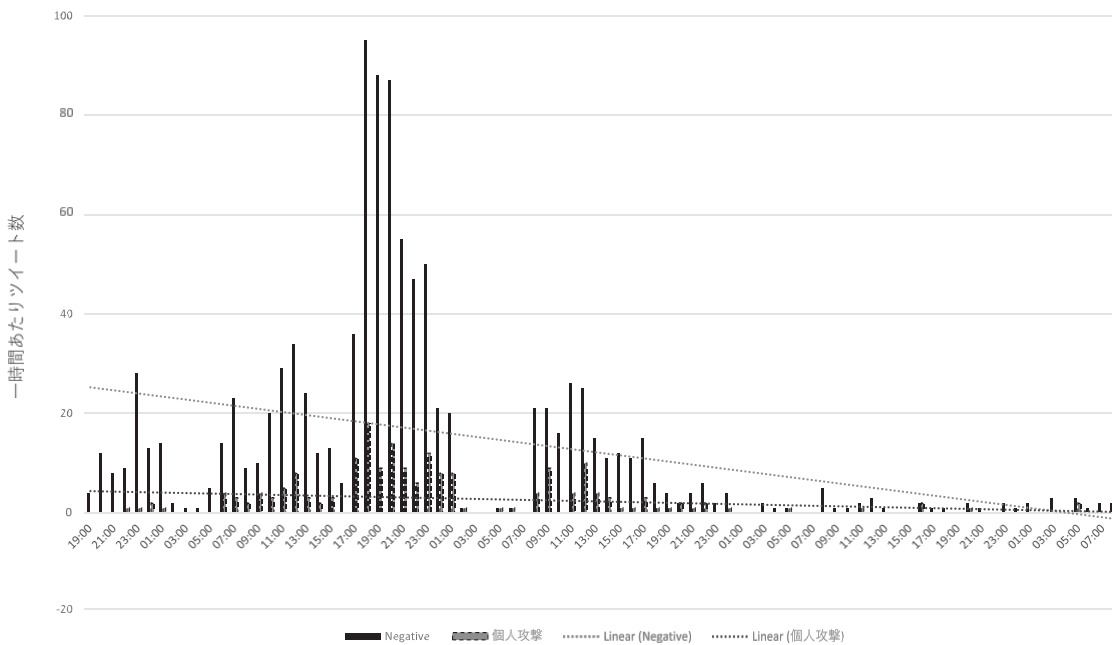

Figure 10. critical and personal attack comments per hour.

The "Twitter account information" of Mr. Z and several other reporters related to Company X is posted in the "Twitter account information" section of this report. Further investigation and analysis are needed to determine whether these users repeatedly troll Mr. Z's other tweets, and what motives they have for their trolling.[8]

---

[8] For the principles of collective action of the troll community, which is connected by a loose network on the Internet, see Ebner (2021). Trolling has existed since the dawn of the Internet, but since the 2016 U.S. presidential election, political warfare in the form of trolling has become the new normal, particularly targeting female artists, activists, and politicians (p. 162,
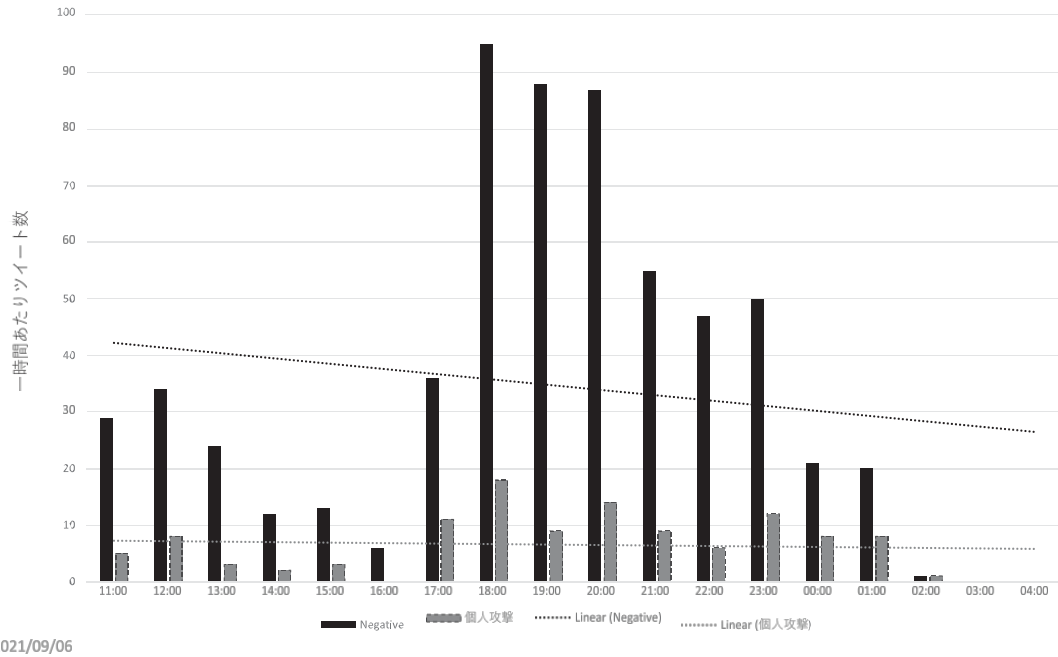
p. 181).

2021/09/06

Figure 11. critical and personal attack comments per hour.
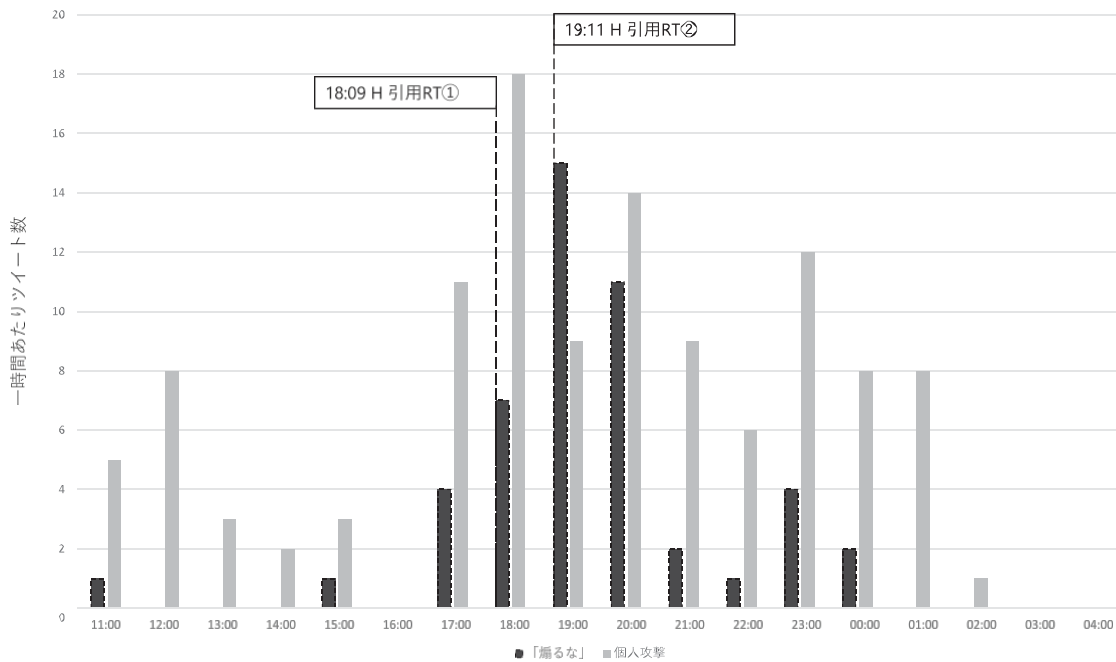

2021/09/06

Figure 12. Comparison of "Don't stir the pot" comments and personal attack comments.

Online Harassment Prevention

Existing studies in Japan and abroad have revealed that anyone can be subjected to online harassment, that on the other hand, there are attributes that make one susceptible to harassment, and that the damage caused by harassment is enormous, both physically and emotionally. In addition, as a case study of Twitter in a Japanese environment, we analyzed the "flame" incident of a reporter tweet from Company X, and found that there were three layers of users who engaged in online harassment: the influencer group, the group of users who were involved in flame wars in response to the influencer's dog whistle, and the group of users who were trolling,

each in the form of a group of users who became The results showed that each of these three groups of users were harassing a single user in a particular way. In response to such online harassment,

What measures can users take?

Yankovic et al. (2021), who conducted a survey and analysis of online harassment of female legislators in various countries, report
Policy recommendations are described. Specifically, the recommendations are for SNS platform companies,

- Introduction of incident reporting system instead of reporting individual tweets

- Sharing of information from the platform to the SNS staff of each company/organization

- Ensuring that the Terms of Use are applied

- Enhanced nudge function

- Establishment of an online harassment prevention mechanism between platforms

As a recommendation to policy makers,

- Tighter regulation of social networking sites (transparent reporting systems, reporting requirements for platforms)

- In addition to deep fakes, etc., gendered harassment should be included in the regulatory scheme.

- The Violence Against Women Prevention Act of 1994 in the United States should include online harm as well as physical harm (Congress.gov, 1994)

As a recommendation to employers,

- Establish guidelines for employees' social networking outgoing communications.

- Psychological, moral and financial support. Payment of service fees for measures to prevent exposure of victims of online harassment, and reporting damage to the platform on their behalf.

The following is a list of some of the most common types of harassment. However, in the case of the Japanese Internet environment, social awareness of online harassment is low and the measures taken by platform companies, which are mainly headquartered in foreign countries, are not sufficient, so harassment prevention measures are required to be taken by individual users or by organizations to which individuals belong, which is the The following is a brief overview of the peculiarities of the Japanese-language SNS environment. The following section discusses measures that individuals and organizations can take to prevent online harassment, with a focus on Twitter, while keeping in mind the particularities of the Japanese-language social networking environment.[9]

Online Harassment Prevention (Personal)

The three most basic anti-harassment actions are muting (a Twitter feature that prevents certain accounts' tweets from appearing on the timeline) blocking (prohibiting certain accounts from sending messages, viewing tweets, or following) and reporting (reporting offensive tweets or (reporting offensive tweets or accounts to Twitter management) (PEN America, n.d.-d). It is also important to keep a log of when and where the harassment occurred. Such information is important evidence when discussing online harassment behavior with the police or an attorney (Table 2).

Recently, dedicated tools for online harassment prevention have also been developed, and the introduction of such tools can also be an indirect form of harassment prevention. For example, a tool called Block Party allows users to set a filter to determine who they want to see replies from and who they do not want to see replies from, and automatically mutes all replies that meet the criteria. This

---

[9] This paper focuses on the analysis of harassment on Twitter, but future studies of anti-harassment measures on other social networking sites will be conducted in the future.
  We would like to make it an issue.

Table 2: Harassment Response Log.

Date	Hours.	Details	Result / What to do ti

This would allow users to check the content of replies on their own at a later date or by asking someone they trust to do so, without having to open Twitter and immediately see the harassing tweet (Cho, n.d.). Additionally, it was announced that Twitter is considering a "safe mode" feature (Doherty, 2021). When enabled, this feature will automatically block users who use harmful language against them or troll for seven days, according to the company. Other possible measures include general digital hygiene checks and team operation of social networking accounts (Jankowicz et al., 2021).

Online Harassment Prevention (Organization)

Nowadays, even those who have traditionally communicated through existing media, such as legislators, reporters, and researchers, are required to communicate through SNS and widely outreach to society. In the case of Mr. Z, harassment took place because of his affiliation with Company X. In the case of Mr. Z, harassment took place because of his affiliation with Company X. Flaming and online harassment have occurred not only against individual accounts, but also against organizations of various sizes (Sakaki and Toriumi, 2020). Therefore, it is difficult for an organization to be completely uninvolved in the operation of social networking accounts of its members and the associated harassment damages.

As a countermeasure that organizations can take, it is essential to first understand the flame-out settlement mechanism. Tashiro and Orita (2012) regarded online flame-outs as an attempt to settle a "glitch," and that a "settlement" is necessary for a flame-out to be resolved. In order for harassment to be resolved, the attacker must be satisfied, and if there is a legal problem, it must be dealt with legally in order to bring the situation to a conclusion (Figure 13). Even if an apology is given, convergence does not occur simply by apologizing; convergence occurs when the apology is accepted. Flames may converge by leading to discussion, by deleting the comment section or the blog itself (in the case of Twitter, by restricting the reply function or deleting the Twitter account), or by ignoring the flames and forgetting about it. In the unlikely event that a member's social networking account is flamed or he/she becomes a victim of harassment, he/she will be criticized and punished for such behavior, so
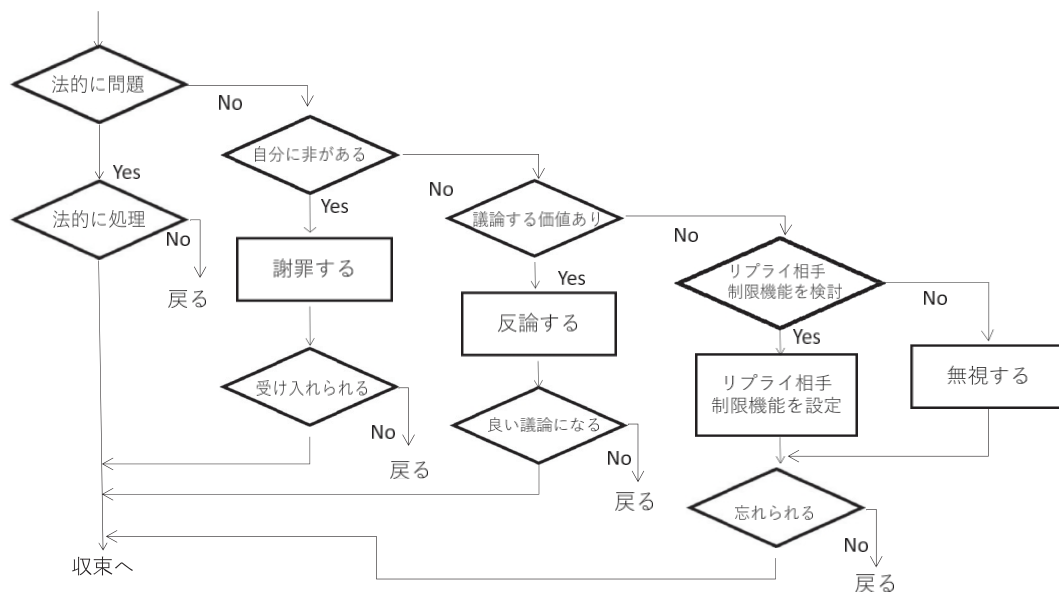
Figure 13: Settlement flow (assuming Twitter, based on Tashiro and Orita 2012).

The first step in supporting SNS operators would be to understand that there are various patterns of convergence. This is a qualitative study, not an empirical or hypothesis testing one, and the question of this paper is a descriptive one, "What is going on? and not a prescriptive question, "What should we do? (Otani, 2019), we would be cautious about using the findings from the analysis in this paper as an aid to countermeasures. On the other hand, it may be useful to interpret the phenomenon of flaming and harassment by keeping in mind that the cases treated in the previous section allowed us to observe that there were several strata of users who engaged in online harassment, each of which harassed in different ways.

Second, it is necessary to identify troll users based on what attributes the organization is given on the Internet. In the case study analysis, the example of media company X was treated, but other examples include the case of Kao Corporation, a major consumer goods chemical manufacturer, which was criticized on the Internet for spreading misinformation that it was a company that contributed to Korea and destroyed Japanese culture because it sponsored many Fuji TV programs that often played Korean dramas (Tashiro and Orita, 2012). As mentioned above, online harassment is a common form of harassment. As mentioned earlier, online harassment has a set of attributes that make the perpetrator an easy target, and the probability of being harassed is highest in the common part of the set. Similar to the way companies examine their brand awareness, they may also need to consider the likelihood of being targeted for harassment.

Earlier, we noted that mute, block, and report are the most basic and effective harassment countermeasures for individuals. On the other hand, there are cases where the act of blocking is considered undesirable, such as accounts by members of the Diet, reporters, and members of public organizations, when the information being transmitted is considered public in nature (Nihon Keizai Shimbun, 2021b) Therefore, it is necessary to discuss the use of the block function within the organization, establish guidelines, and operate the function appropriately. If the organization prohibits effective measures that members can take for self-protection, the possibility of physical and psychological damage from harassment will increase, and more substantial psychological, moral, and financial support will be needed.

In Japan, on April 21, 2021, an amendment to the Provider Liability Limitation Law was passed to make it easier to identify people who post defamatory statements on the Internet, simplifying the procedures for identifying perpetrators. On October 7, it was also reported that a draft outline of a legislative amendment to the Penal Code adding a penalty of imprisonment for "insult" was compiled by a subcommittee of the Legislative Council, which will report to the Minister of Justice (Kyodo News, 2021). Nevertheless, the burden of direct and indirect costs of online harassment, such as disclosure procedures, remains heavier for victims than for perpetrators. Therefore, organizations should consider providing financial support to their members, for example, in the form of partial coverage of the cost of lawyer insurance.

Conclusion

Previous studies in Japan and abroad have revealed that while anyone can be subjected to online harassment, there are certain attributes that make one susceptible to harassment, and that the physical and psychological damage suffered by victims cannot be overlooked, even if the harassment takes place online. As a case study of online harassment on Twitter in a Japanese-language environment, we analyzed a "flame" incident involving a reporter tweet from a major media company, X. We found that there are three types of users who engage in harassment: the influencer group, the group of users who participate in flame wars in response to the influencer's dog-whistle, and the group of users who troll the Twitter feeds. In addition, we observed that each of them harassed a single user in different ways. Although the case analysis is difficult to generalize since it is only one example of a flaming case, it complements the conclusions of previous studies on the phenomenon of flaming on SNS. A larger-scale analysis based on the assumptions derived in this paper should be the subject of future work.

At the end of this paper, we will conclude by detailing the measures that individual SNS users or their organizations should take to prevent harassment, given the low social awareness of online harassment in the Japanese Internet environment and the difficulty of asking foreign-based platform companies such as Twitter to take measures to prevent such harassment. The report details measures to prevent harassment that should be taken by individual SNS users or the organizations to

which they belong. With regard to SNS in the Japanese language environment, the concept of online harassment itself is not well known, and compared to its social impact, there have been few studies focusing on victims who have been harassed. In this sense, we hope that this paper will help to broaden and deepen academic discussions and lead to effective harassment prevention measures.

data availability

All data underlying the research results of this paper are presented in this paper, and no additional source data is required. An exception is the sharing of the underlying data for this study due to ethical and copyright restrictions on social media data. The data covered in this study are from the August 3, 2021 posting of the large

The tweets of a female job working for a media company (hereinafter referred to as "Ms. Z") and 2,817 direct replies (1,175 direct replies, 1,499 of these retweets, 143 quoted retweets)sent directly to Ms. Z's account between August 3 ~7, 2021, and 5,595 quoted retweets of the main retweets. Detailed information that allows for the replication of this study can be found in the "Case Analysis" section. If you have any questions about the methodology, please contact the authors (tonami.aki.ka@u.tsukuba.ac.jp).

Tweet IDs will also not be shared due to security considerations to protect Mr. Z's anonymity. Researchers in the same field who wish to access the ID with good intentions and clear research objectives should contact the author (tonami.aki.ka@u.tsukuba.ac.jp), stating how and for what purpose the data will be used.

thanks

References

Berry E: Online violence just as destructive as offline violence. 2019. (Accessed 24 September 2021).
Reference Source

BIGLOBE K.K.: More than 80% of respondents think that penalties for slander on social networking sites "should be strengthened. BIGLOBE ti Announces Second "with Corona Era Stress Survey". ~with more than 30% of respondents "feeling" resistance to posting their activities on social networking sites in the age of Corona~. 2020. (accessed 13 September 2021).
Reference Source

Cho T: Japanese usage guide ti Block Party is now available! n.d. (accessed 13 September 2021).
Reference Source

Congress.gov: H.R.3355 - Violent Crime Control and Law Enforcement Act of 1994. 1994. (accessed 20 October 2021).
Reference Source

Correll SJ, Weisshaar KR, Wynn AT, *et al.*: Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment *Am. Sociol. Rev.* 2020; 85(6): 1022-1050.
Publisher Full Text

Dohe rty J: Introducing Safety Mode. 2021. (accessed 13 September 2021).
Reference Source

Fung A, Gilman HR, Shkabatur J: Six Models for the Internet + Politics. *International Studies Review.* 2013; 15(1): 30-47.
Publisher Full Text

Glick P, Fiske ST: The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *J. Pers. Soc. Psychol.* 1996; 70: 491-.
Publisher Full Text

Glick P, Fiske ST: Hostile and Benevolent Sexism. *Psychol. Women Q.* 1997;.
21(1): 119-135.
Publisher Full Text

Glick P, Fiske ST: The Ambivalence toward Men Inventory: Differentiating hostile and benevolent beliefs about men. *Psychol. Women Q.* 1999; 23(3): 519 -536.
Publisher Full Text

Gosse C, Veletsianos G, Hodson J, *et al*.: The hidden costs of connectivity: nature and effects of scholars' online harassment. *Learning Media and Technology.* 2021; 46(3): 264-280.
Publisher Full Text

Inter-Parliamentary Union: *sexism, harassment and violence against women in parliaments in Europe.* Inter-Parliamentary Union; 2018.
Reference Source

Jankowicz N, Hunchak J, Pavluic A, *et al*: *Malign Creativity: How gender, sex and lies are weaponized against women onilne.* Washington D.C.: Wilson Center; 2021.
Reference Source

Khan I: *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinin and expression.* United Nations General Assembly; 2021 July.
Reference Source

Lee E: Into 2020: The State of Online Harassment and Opportunities for Collaboration. *Online SOS.* 2019.
Reference Source

McCay-Peet L, Quan-Haase A: What is social media and what questions can social media research help us answer? Sloan L, Quan-Haase A, editors. *the SAGE Handbook of Social Media Research Methods.* London: SAGE; 2017.

PEN America: Dog Whistling. n.d.-a. (accessed 13 September 2021a).
Reference Source

PEN America: Online Harassment Survey: Key Findings. n.d.-b. (accessed 7 September 2021b).
Reference Source

PEN America: Online Harassment. n.d.-c. (accessed 7 September 2021c).
Reference Source

PEN America: Respond. N.d.-d. (accessed 13 September 2021d).
Reference Source

Sobieraj S: *Credible Threat: Attacks Against Women Online and the Future of Democracy.* Oxford: Oxford University Press; 2020.

Statista Research Department: Leading countries based on number of Twitter users as of July 2021.(accessed 7 October 2021).
Reference Source

Vogels EA: *The State of Online Harassment.* Pew Research Center; 13 January 2021.
Reference Source

Weigard A, Loviska AM, Beltz A: Little evidence for sex or ovarian hormone influences on affective variability. *Sci. Rep.* 2021; 11: 20925. Publisher Full Publisher Full Text

Ybarra ML, Mitchell KJ, Wolak J, *et al*: Examining Characteristics and Associated Distress Related to Internet Harassment: Findings From the Second *Pediatrics.* 2006; 118(4): e1169-e1177.
Publisher Full Text |Reference Source

Ohtani, Hisashi: Concept of qualitative research: From research methodology to analysis by SCAT. Nagoya Univ. Academic Press; 2019.

Kyodo News: "Prison sentence for insult" to be made stricter. 2021. (accessed 7 October 2021).
Reference Source

Koyama, K., Asatani, K., Sakaki, G., *et al.*: Resonance structure of users in online flame wars. *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence.* 2019.

Takeshi Sakaki and Fujio Toriumi: An analysis of incitement of media company SNS accounts based on the social pornography hypothesis. *The 34th Annual Conference of the Japanese Society for Artificial Intelligence.* 2020.

Shang Tisong, Zhou Juan, Takada Hsi: Online harassment countermeasure by human relations and slander detection. FIT2018 (17th Forum on Information Science and Technology). 2018.
Reference Source

Sanseido: Super Daijirin. 2013.

Shin, Qi-young and Mari Hamada: Online Harassment of Female Council Members - Metropolitan Area.
The case of a female local councilor in the United States of America. Livelihood and Economic Policy. 2021; 291: 291.
19-23.

Mitsuteru Tashiro and Akiko Orita: A Study on the Process of Outbreak and Convergence of Internet Flames.
In Proceedings of the 9th Annual Meeting of the Information Processing Society of Japan. IPSJ Research Report. 2012.

TANAKA, Toko: Why are women more likely to be flamed on the Internet than men?

'I want to punish women who don't act like women'. 2020.(accessed 22 December 2021).
Reference Source

Tomoko Nagasako: "Disinformation" in Today's World

-- "Fake News" t i e t a l . to "Disinformation". February 2021 (accessed 22 December 2021).
Reference Source

Nihon Keizai Shimbun: Tokyo, 5,386 new infections, second highest on record, 7-day average for previous week

117.9% of the 2021a.(accessed 20 October 2021).
Reference Source

Nihon Keizai Shimbun: Kono's Twitter account blocked. Nihon Keizai Shimbun

The Daily Yomiuri. 8 September 2021b.
Reference Source

Plan International: Findings on Gender-Based Online Harassment of Young Women Japan Survey Report. 2020 October.
Reference Source

Mariko Morishita: Mobile Internet use ti increase at home! Information media needs highlighted by the Corona disaster ti. 2021. (accessed 22 September 2021).
Reference Source

Morinaga, Y.: Gender Stereotypes. In Ryoko Kimura, Kumiko Ida, and Kimie Kumayasu (eds): Understanding Gender Studies in the Humanities, Social Sciences, and Natural Sciences. Minerva Shobo. 2013; 96-97.

Yamaguchi, S.: Verification of the Actual Conditions of Flames and Attributes of Flamers by Empirical Analysis. Information and Communication Studies

Journal of the Society. 2015; 33(2): 53-65.
Publisher Full Text

Yamaguchi, S.: An Empirical Analysis of Motives for Engaging in Flaming.

2016 Society for Socio-Informatics Studies (SSI) Conference. 2016.
Reference Source

Yamaguchi, Shinichi: Identity of "extreme people" who wield justice. Kobunsha; 2020.

Julia Ebner: Going Dark: 12 Undercover Reports on 12 Extremist Organizations. Left Right Publishing.
2021.

Open Peer Review

Current Peer Review Status: Current Peer Review Status ✓ ✓

Version 2

Reviewer Report 21 February 2022

https://doi.org/10.5256/f1000research.120282.r122066

✓ **Shinichi Yamaguchi** [iD]

Center for Global Communications, International University of Japan, Niigata, Japan

The paper was an excellent one with high social significance from the time of submission, and all appropriate corrections and rebuttals were made in response to the review comments. The improvement in various aspects, such as the enrichment of the review of previous research, the display of the research question, the explanation of the methods, and the display of the research problem, was seen, and I believe that the manuscript has reached an extremely high level of perfection without being considered unfamiliar with the language.

Despite the fact that online harassment is an important issue with a large number of victims, there are few studies, especially in Japan. We hope that this paper, which shows the reality of online harassment and its countermeasures, will help many individuals and organizations, and contribute to the improvement of services and the creation of appropriate rules.

On the other hand, as pointed out by the authors and another reviewer, Mr. Kawabata, there are several issues to be addressed, and I believe that this is a topic that can still be expanded from various perspectives. I am very much looking forward to the future development of this research.

Competing Interests*:* No competing interests were disclosed.

Reviewer Expertise: Econometrics, Social Media, Social Informatics, Information Economics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 20 December 2021

https://doi.org/10.5256/f1000research.78433.r100521

? **Shinichi Yamaguchi** (iD)

1 Center for Global Communications, International University of Japan, Niigata, Japan

2 Center for Global Communications, International University of Japan, Niigata, Japan

This study introduces the concept of online harassment, clarifies its outline, and analyzes flaming cases on Twitter to identify users who engage in harassment. In the last part of the paper, measures that should be taken by individuals and organizations are described in detail. As the authors point out, there have been few studies that analyze flame wars from the perspective of the victims, and it is significant that the authors have introduced the concept of online harassment in their analysis. The empirical analysis is also very detailed, and the field is of great interest to the public, and many readers may be interested in the research results.

However, I think it has some problems when considered as an academic paper. Specifically, I think the following points could be improved

(1) Review of existing studies - analysis - no relevance of measures

I think the biggest thing is the lack of connection between the sections and the report-like listing of opinions (each

(The opinion itself is unoptimistic, but as an academic document, it has its problems.) For example, the lengthy article (Fig. 2) points out that "there are mechanisms by which malicious actors harm their targets (victims) through a variety of means, in a number of media, in a number of locations", but only Twitter is analyzed. Furthermore, "Online Harassment Countermeasures," which devotes a considerable amount of space to the analysis, contains a long list of descriptions that have absolutely nothing to do with the empirical analysis. The patterns of harassers are not utilized.

One of the factors that may contribute to this is the structure of the section. First, in the beginning, the "Introduction" and "Online - Haras

The latter can be said to be a review of existing research. The latter is a review of existing research, but it focuses on describing the current situation and does not clearly state what issues exist in the existing research, and therefore, does not clearly state the research question that this research will clarify. If I had to use a strong language, I would say that there are only studies in European and foreign language environments, so I would say that there is a need to clarify the issues in the Japanese language environment.

I think that "describing a case study" is the right place to do so, but I think it would be better to clarify the position of this study in relation to existing studies and the Research Question in order to show the novelty and social significance of this study. For this purpose, I think it is necessary to have something more relevant to the content of this empirical analysis than the existing studies. It may be better to have a separate section on the Research Gap and Research Question.

Also, in the case analysis section, the method and the analysis results are placed in the same section, so write them separately, and also include the method in the section.

It would be better to be more detailed. In particular, about the classification of the three tiers, which is a key finding in this study,

The only thing it says is "Analyze the relationship between the original tweets and the pull RTs," and I couldn't understand the method.

It would be better to include a section discussing the relationship between the results and previous studies, and a concluding section discussing countermeasures derived from the results. Although the contents of the current "Online Harassment Prevention" are interesting, we recommend reducing the size of these sections and devoting more space to the discussion of countermeasures that can be derived from the analysis.

(2) There is only one case study.

I thought it was a bit unreasonable to analyze online harassment patterns based on a single case study. Since this is a case study analysis, the number of cases does not need to be large, but it

would be better to have a few more cases to verify the robustness. However, I think this would be difficult, so this is not a request for modification, but a proposal for modification.

Other (minor points)

-In the analysis, there is a question whether "don't get emotional" alone can be interpreted as racist by those who live in the area where the woman lives. I understand that it is often used in a discriminatory manner, but it is a language that is also used for male partners. Also, if "using "sincere disrespect" is an attack on the person, then "don't be emotional" is also an attack on the person.

-I couldn't find ЈКоЩ (2019) in the literature list.

Is the research clearly and accurately presented and does it draw on the latest literature?

Yes, sir.

Is the research design appropriate and of academic value?

Kogei Partially Applicable

Are sufficient details of the methodology and analysis provided to enable reproduction by a third party?

Kogei Partially Applicable

(Response required if applicable) Are the statistical analyses and interpretations adequate?

Kogei Partially Applicable

Are all source data that will serve as the basis for the results available and reproducibility guaranteed to the fullest extent?

No source data required

Are the conclusions reasonably supported by the results?

Kogei Partially Applicable

Competing Interests*:* No competing interests were disclosed.

Reviewer Expertise: Econometrics, Social Media, Social Informatics, Information Economics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 13 Jan 2022

Aki Tonami

Mr. Yamaguchi

Thank you very much for your peer review comments.

(1) Review of existing studies - analysis - no relevance of measures

The premise of the journal's research articles is that "Research Articles should present originality in findings and insights and offer theoretical, empirical, experimental and/or Null and negative findings and reanalyses of previous studies leading to new results, as well as as as new research in their respective fields of research. Null and negative findings and reanalyses of previous studies leading to new results, as well as confirmatory results, are also encouraged" (https://f1000research.com/for-authors/article-guidelines). The degree to which one should devote space to the review of previous studies in order to present an authenticity of the research depends on the field, but in this paper we have chosen to focus on the following three aspects of online harassment, since it is highly dependent on language, culture and context

The report avoids an exhaustive review of previous research on online harassment in the English-speaking world, primarily in English, and discusses online harassment in the Japanese-language environment, where it does not exist in the first place. We aimed to become the first company to be considered. Therefore,

The research gap is the absence of existing research. This is a qualitative study and is not intended to be an empirical study or to test a hypothesis. Therefore, the research question should be "online

-The question is descriptive: "What is the current state of harassment? However, we understand that you may have pointed out that this is difficult to understand, so we have revised the question to that effect.

○ For a Japanese-language study on online harassment published after the publication of this article, see, for example, the special issue of the Law Seminar December 2021, entitled "Forgiveness and Sanctions against Language".

○ As for your comment that "there are issues with the academic texts," we would appreciate it if you could clarify the assumptions and conditions you are looking for, as it is unclear what exactly you are referring to in your academic texts. (For example, is it a quantitative research paper, which is often written in IMRaD format?
)

○ In Figure 2, online harassment spans multiple platforms other than Twitter.
It is correct to point out that this is being carried out. However, since harassment on multiple platforms is outside the scope of this report, we have added a footnote as a future issue.

○ The method "should be written in more detail" and "analyze the relationship between the RTs of the original tweet pulls.
I have described the details of the "What can I do? If you have any other shortcomings, please let us know what you think is appropriate and how much detail you need to provide.

(2) Comment: "There is only one case study."
As stated in the article, it is difficult to generalize only from this case and it complements the conclusions of previous studies on the phenomenon of flame wars on social networking sites.

Other (minor points)
○ I think your first comment is about coding when performing qualitative data analysis.
However, "women are emotional" is a gender stereotype ("女 is kind").
The personalities and abilities of men and women, such as "men are strong," "women are caring," and "men are good with numbers.
Morinaga, 2013), and these prejudices distort perceptions of others and lead to sexism, especially benevolent sexism (Glick & Fiske, 1996, 1999, Correll et al., 2020).
Glick & Fiske, 1996, 1997, 1999, Correll et al., 2020).
By attacking the performance of the utterance as "emotional" to the female narrator, the intent of the speaker is
It has been pointed out that they suppress their opinions (solder, 2020). You point out that men are also sometimes criticized for being "emotional" - as mentioned above,
It is believed that the stereotypes that are generally associated with the character, ability, role, and behavior of the families are pointed out and criticized because they are perceived as incorrect and unfavorable. For these reasons, we have determined that "emotional" is not a personal attack, but a discriminatory term against the group to which the subject belongs (or is considered to belong).

○ On the other hand, we have placed attacks on individuals in this category, not the group to which the subject belongs (in this case, the women's gender group or the media company), but the reporter himself/herself. We have added a clear example of this, while keeping in mind the protection of copyrights and privacy.

○ In this paper, the coding is done in an exploratory manner while reading the data.
Generative coding" was conducted. In this type of coding, which is often used in qualitative research, interpretation plays an important function and is considered to be

proactive. For this reason, it is rare to have more than one coder, and this point is not considered to be a problem with the "generative coding" method, in which codes are selected from a set of selected codes and assigned to multiple coders to evaluate the degree of agreement between them.

The term "standardized coding" is not the same as "standardized coding" (大谷谷 2019).

○ With respect to online harassment prevention, as noted earlier, this paper is a qualitative study, with no empirical or
The purpose is not to test hypotheses, and the research question is not intended to be a "what is going on" question.
?" (Ōtsubo 2019), and not "what should we do?" (Ōtsubo 2019). (Ō谷2019), we have avoided using the findings from this analysis as a guide to countermeasures. On the other hand, previous studies, mainly in English-speaking countries, and previous Japanese-language studies on "flames" have provided valuable opinions that should be used as a countermeasure. We have reviewed them from the practical point of view of the Japanese-language Internet environment and the organizations to which you belong, and we have included what we considered useful. We believe that it is as you pointed out that this background is difficult to understand, so we have made some corrections.

○ Thank you for pointing out the omission in the bibliography of the article by Kohara et al. It has been inserted along with other additional references.

references

Correll SJ, Weisshaar KR, Wynn AT, *et al.*: Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment *Am. Sociol. Rev.* 2020; 85(6): 1022-1050. https://doi.org/10.1177/0003122420962080

Glick P, Fiske ST: The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *J. Pers. Soc. Psychol.* 1996; 70: 491- 512. https://doi.org/10.1037/0022-3514.70.3.491

Glick P, Fiske ST: Hostile and Benevolent Sexism. *Psychol. Women Q.* 1997; 21(1): 119-135. https://doi.org/10.1111/j.1471-6402.1997 tb00104.x

Glick P, Fiske ST: The Ambivalence toward Men Inventory: Differentiating hostile and benevolent beliefs about men. *Psychol. Women Q.* 1999; 23(3): 519 -536. https://doi.org/10.1111/j.1471-6402.1999.tb00379.x

Weigard A, Loviska AM, Beltz A: Little evidence for sex or ovarian hormone influences on affective variability. *Scientific Reports*. 2021; 11, 20925. https://doi.org/10.1038/s41598-021-00143-7

T. Oyamata: Directions for qualitative research: From methodology to SCAT analysis. Nagoya University Press; 2019.

D. Solanaka: Why are women more likely to be burned on the Internet than men? Reference Source

Kohhei Kohara, Kimitake Asa, Takeshi Sakaki, *et al.*: Resonance structure of users in online flame wars. *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence.* 2019.

Competing Interests***:*** No competing interests were disclosed.

Reviewer Report 09 December 2021

✔   **Yasuko Kawahata**

[1] College of Social Relations, Department of Communication and Media Studies, Rikkyo University Tokyo, Japan
[2] College of Social Relations, Department of Communication and Media Studies, Rikkyo University Tokyo, Japan

Introduction - The preamble contains a discussion of research, case studies, etc. on the latest developments in online harassment and its origins, and we thought it would serve as collective information to summarize the current online risk management discussion for the general public. We thought it would serve as a collective information that would summarize the current online risk management debates for the general public.

（1）Overall Review

In this case, I think that this article is a comparatively remarkable case of criminals targeting journalists.

At the same time, the discussion on the issue of online-ha is a very important one, which will continue to be discussed in the future.

I believe that this book also plays a role as a text and document that displays a way to solve the problem of harassment. In this discussion, we developed a discussion based on the example of a journalist (Mr. Z), but it is also an issue that is appearing in a wide range of media units and is also a social issue from a micro perspective and within the general community.

-We believe that this is an issue that needs to be discussed in the future because there is a risk of online use as a form of harassment. In this paper, we have considered the privacy of both the victim and the offender with regard to the source data on SNS that was the subject of the analysis (approved by the editorial department), and we hope that this will contribute to deepening the discussion. We hope that this will contribute to deepening the discussion.

（2）Points to follow up on

As a topic related to this paper, it is an issue how the revised policy on impersonation on Twitter - Terms of Use, which was released on November 31, 2021, will be handled in this discussion and whether it can work effectively.

"Twitter Safety" states that "there is a potential for undue influence among families, activists, dissidents, and members of minority communities," and that if an image or video is shared that violates the policy, Twitter will remove the media and take appropriate action. If an image or video is shared that violates the policy, Twitter will remove the media and take appropriate enforcement actions, such as making the tweet less visible in search results, deleting the tweet and notifying the poster, and Twitter has the authority to permanently freeze user accounts that violate the policy. (Intended Translation - Adapted from: https://twitter.com/TwitterSafety, Ref: 2021/12/02)

We believe that the discussion in this paper will be further accelerated as cases related to specific measures such as the following will also be discussed.

As suggested in this paper, the main reason for this is to "impersonate" a "specific person" who performs aggressive acts on an unspecified number of people.

There have also been cases of shady social harassment, or so-called "social pornography," such as "implicitly" defaming a person after "getting rid of" him or her.

In addition, unauthorized media exploitation, ridicule of nasal characteristics, antisocial acts, and other acts of violence against the target of the exploitation are prohibited.

Although the environment has started to improve on a platform basis in terms of information sharing and warnings such as the above-mentioned reference, there are still many elements that need to be worked on to solve problems.

(3) Future issues related to this discussion and research prospects

The challenge in this paper is that online harassment has shifted from texting and other forms of social harassment, such as short messages, to the exposure of social pornography in the form of images, illustrations and videos, which are more difficult to track and trace. This is also the case for the case studies in this paper. Therefore, we believe that more attention should be paid to cases such as the one described in this article, and that information should be shared with educational institutions to alert them to the situation. Especially with the use of highly diffuse media such as images, audio, and video, the risk of social pornography is even higher, and the victims are more likely to be victims of social pornography.

The fact-checking around the perpetrator can be difficult, and cover-ups can easily be made.

The case of entering the house has also been observed as a victim's voice on social networking sites. Therefore, we believe that an opportunity to share awareness of the problem is necessary first. (In the above case, the opinion was based on an anonymous survey by the reviewer.) In this regard, we believe that the contents of this report have academic value. This is an issue that needs to be addressed in the future, and we should start discussing the overheated discussions on SNS when solving these problems, reasonable consideration for perpetrators and victims, and dispute resolution, etc., to prevent illegal activities on the Internet.

-In addition, we hope to develop a more academic discussion in light of the revision of the Act on the Limitation of Liability of Providers for Harmful Information (Provider Liability Limitation Act) in this fiscal year. In addition, we believe that in developing this discussion, it is necessary to take into account the risk of sudden exposure to harmful information, such as "risk management" on SNS and information sharing among those involved in education. On the other hand, I believe that the discussions and documents in this paper will be useful for information sharing.

(4) summary

As mentioned in this paper, we believe that it is also necessary to collect case studies of incidents where we are presenting suggestions regarding harassment-preventive measures.

Especially for those who are vulnerable to exploitation, such as the young and the vulnerable who do not have the knowledge or the ability to bring a lawsuit or other legal action.

Incidents will continue to occur in the future, so there is a need to collect information on these issues, discuss countermeasures, share information, and promote community building. As mentioned in the article and documents, it often takes time to resolve incidents and disputes, especially on platforms located in overseas servers, and there should be continuous attempts to discuss effective preventive measures and opportunities to share opinions in journalism such as articles and documents. I hope that this article will help to deepen the discussion.

references

Twitter, Inc.: Impersonation Policy. n.d. Available at: https://help.twitter.com/ja/rules-and-policies/twitter-impersonation-policy. Accessed 2021/12/02.

Ministry of Internal Affairs and Communications: Response to illegal and harmful information on the Internet (Provider Liability Limitation Act). n.d. available at:: MIC: Ministry of Internal Affairs and Communications.

https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/ihoyugai.html

Is the research clearly and accurately presented and does it draw on the latest literature?

Yes, sir.

Is the research design appropriate and of academic value?

Yes, sir.

Are sufficient details of the methodology and analysis provided to enable reproduction by a third party?

Yes, sir.

(Response required if applicable) Are the statistical analyses and interpretations adequate?

Not applicable (not using statistics)

Are all source data that will serve as the basis for the results available and reproducibility guaranteed to the fullest extent?

No source data required

Are the conclusions reasonably supported by the results?

Yes, sir.

Competing Interests*:* No competing interests were disclosed.

Reviewer Expertise: Social Networking, Social Information, Media Society

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 13 Jan 2022

Aki Tonami

Kawabata, a resident of Kawabata

Thank you very much for your peer review comments.
We believe that all of the comments we received are very important and will be helpful for the development of future research. Regarding points (2) and (3), we have added the following as future issues: the need to consider cross-platform harassment, identity theft, and harassment using other means than text.

Competing Interests*:* No competing interests were disclosed.