Online harassment of Japanese celebrities and influencers

Masanori Takano[1]*, Fumiaki Taka[2], Chiki Ogiue[3] and
Natsuki Nagata[4]

[1]Multidisciplinary Information Science Center, CyberAgent, Inc., Tokyo, Japan, [2]Faculty of Sociology,
Toyo University, Tokyo, Japan, [3]Chiki Lab, Tokyo, Japan, [4]Graduate School of Education, Hyogo
University of Teacher Education, Hyogo, Japan

Famous people, such as celebrities and influencers, are harassed online on a
daily basis. Online harassment mentally disturbs them and negatively affects
society. However, limited studies have been conducted on the online harassment
victimization of famous people, and its effects remain unclear. We surveyed
Japanese famous people ($N = 213$), who were influential people who appeared
on television and other traditional media and on social media, regarding
online harassment victimization, emotional injury, and action against offenders
and revealed that various forms of online harassment are prevalent. Some
victims used the anti-harassment functions provided by weblogs and social
media systems (e.g., blocking/muting/reporting offender accounts and closing
comment forms), talked about their victimization to close people, and contacted
relevant authorities to take legal action (talent agencies, legal consultants, and
police). By contrast, some victims felt compelled to accept harassment and
did not initiate action for offenses. We propose several approaches to support
victims, inhibit online harassment, and educate people. Our findings help that
platforms establish support systems against online harassment.

KEYWORDS

online harassment, celebrity, influencer, online questionnaire survey, social media
platform, support systems against online harassment

1 Introduction

Online harassment is a critical issue resulting from Internet use and considerably
harms victims' mental health and dignity. Online harassment includes cyberbullying,
hate speech, flaming, doxing, impersonating, and public shaming (Blackwell et al., 2017).
Famous people, such as celebrities and influencers, are harassed online on a daily basis
(Van den Bulck et al., 2014; Ouvrein et al., 2018; Hand, 2021); however, the extent of
injury and actions remains unclear because of the difficulties in conducting investigative
surveys on famous people. Most studies investigating the online harassment of celebrities
are case studies using social media logs, such as Kim et al. (2013), Lawson (2017), Marwick
(2017), Matamoros-Fernández (2017), Lawson (2018), and Park and Kim (2021). Surveys
of victims can provide an understanding of the online harassment of famous people. For
example, surveys of female journalists (Chen et al., 2018), Japanese journalists (Yamaguchi
et al., 2023a), and politicians (Every-Palmer et al., 2015) have revealed victimization, mental
and physical injuries, effects on their day-to-day activities (e.g., involuntarily inhibiting
victims' activities), and actions against online harassment. Our study reveals a quantitative
picture of the online harassment victimization of famous people through a survey ($N =$
213). Accordingly, we discuss the mitigation of online harassment victimization.

The online harassment of famous people can easily become particularly extreme (Lawson, 2017; Ouvrein et al., 2018; Scott et al., 2019; Lee et al., 2020) thereby harming victims (Bliuc et al., 2018) and can lead to suicide (Hinduja and Patchin, 2010; Brailovskaia et al., 2018).

Online harassment against famous people also causes indirect negative effects. First, sensationalized coverage of the suicides of famous people increases suicide mimicry (Werther effect) (Phillips, 1974; Kim et al., 2013) because such coverage evokes increased negative emotions and feelings of social isolation than the coverage of suicides of other victims (Rosen et al., 2019). Second, observing online harassment encourages the aggressive behavior of observers (Bliuc et al., 2018; Ouvrein et al., 2018; Scott et al., 2019; Yokotani and Takano, 2021) and tempts them to justify their offensive behavior (Crandall and Eshleman, 2003), e.g., they believe that celebrities are socially strong (Lawson, 2017; Ouvrein et al., 2018; Lee et al., 2020). Consequently, online harassment would escalate in intensity and frequency. Third, spreading negative information that damages victims' reputations, such as negative gossip and disinformation, severely affects their businesses (Sarna and Bhatia, 2017). Fourth, the popularization of online harassment inhibits free expression and discussions among victims and other people (Chen et al., 2018; Yamaguchi et al., 2023a).

Although famous people can be vulnerable to online harassment, they do not always act against it. Victims may feel stigmatized by harassment and fear rejection by others when they speak out (Mesch and Talmud, 2006; Andalibi et al., 2016; Foster and Fullagar, 2018). A famous victim targeted by several people tends to attract victim-blaming (Hand, 2021). Furthermore, the victimization of famous people tends to attract sensational media coverage (Kang et al., 2022). Therefore, victims are often under severe suppression (Hand, 2021; Kang et al., 2022).

Additionally, the anti-harassment actions taken by victims are often associated with several problems. Offenders typically circumvent blocking/muting accounts by creating new ones. The Internet is a vital tool for communication between famous people, such as celebrities and influencers, and their fans (Van den Bulck et al., 2014). Therefore, closing comment forms decreases communication opportunities with fans and impairs business. Actions against online harassment on other platforms (indirect harassment, e.g., harmful gossiping about victims) are limited (Sarna and Bhatia, 2017) and require complex legal procedures. Therefore, many victims are forced to accept online harassment.

Social media platforms remove toxic comments and ban the accounts of people posting such comments to protect their users, including celebrities and influencers. However, the efficacy of these provisions is limited. Social media platforms cannot moderate all toxic comments (Mohammad et al., 2016; Cho, 2017; Badjatiya et al., 2019; Sap et al., 2019; Milosevic et al., 2022). Linguistic filtering approaches can accidentally result in the biased treatment of discriminated minorities (Badjatiya et al., 2019; Sap et al., 2019). Removing toxic comments according to the terms of service only removes explicit expressions; therefore, ambiguous and/or cloaked blatant expressions tend to remain on platforms (Mohammad et al., 2016; Cho, 2017; Milosevic et al., 2022). A study revealed that banning the accounts of offenders would not alter their thinking (Johnson et al., 2019).

Online harassment victimization has some common characteristics across countries and cultures (Chen et al., 2018). Famous Japanese people are also harassed online; e.g., Japanese journalists were more than four times more likely to experience online harassment than the general population (Yamaguchi et al., 2023a); and Japanese professional wrestler Hana Kimura committed suicide owing to online harassment (Dooley and Hida, 2020). A survey of Japanese Internet users conducted shortly after her suicide showed that the respondents tended to think that online harassment against famous people is the price of fame and/or unavoidable accidents (BIGLOBE Inc, 2020), as also found by previous studies in other countries (Lawson, 2017; Ouvrein et al., 2018; Lee et al., 2020). A survey of Japanese journalists also indicate that the lack of organizational cover makes them easy targets for harassment, and that the number of people who take concrete action (e.g., blocking and talking family and police) when victimized is limited (Yamaguchi et al., 2023a). The lack of such actions has been noted in surveys of ordinary Japanese Internet users (Yamaguchi et al., 2023a). Additionally, the Japanese anonymous bulletin board 2channel, which is a toxic site on the Japanese Internet, has cultural roots in 4chan, which is a toxic site on the English Internet (Dooley and Ueno, 2022). Therefore, studying harmful online phenomena in Japan may have value beyond Japanese culture.

We study the following research questions:

- **RQ1**: What is the relationship between victims' features and harassment types?
- **RQ2**: How do victims react to online harassment?

Based on the answers to these questions, we reveal the issues of the current online harassment victim support system and discuss ways to improve them.

The contributions of the study are as follows:

- A quantitative picture of the online harassment victimization of famous people is revealed.
- The characteristics of vulnerable famous people in online harassment are discussed.
- Issues with existing support systems against online harassment for famous people are identified.
- Based on these, we discuss approaches against online harassment.

2 Materials and methods

2.1 Participants

We conducted an online survey of famous people (September 27 to November 18, 2021). The participants in this survey were recruited by a Japanese Internet company, CyberAgent, Inc.[1] They, who were business partners of CyberAgent, Inc., were influential people who appeared on television and other traditional media and on social media. The number of potential participants was ~20,000. The details of the population cannot be disclosed as they are a trade

---

1   https://www.cyberagent.co.jp/en/

secret of CyberAgent Inc. Participation in this survey is unpaid. We analyzed the results of questionnaires that they fully answered ($N = 213$). The participation rate of this survey was low; this may be because the survey was conducted without monetary compensation for respondents, and celebrities and influencers generally tend to have full schedules. Note that these participants appeared to have participant biases because victims may tend to participate in this survey.

2.2 Measures

Our survey addressed victimization by online harassment. We created questionnaire items by referring to a previous survey of harassment in Japan.[2] Additionally, the participants' responses concerned emotional injury, actions against online harassment, victimization through offline harassment, demographic information, the number of online followers, and media appearance. We investigated associations between them and victimization through online harassment.

2.2.1 Direct and indirect online harassment

Offenders can directly attack victims online (Van den Bulck et al., 2014; Sarna and Bhatia, 2017; Hand, 2021). Online attacks include posting toxic comments to victims' weblogs, sending harmful replies to victims' social media accounts, and sending sexual messages to victims. We investigated victimization through direct online harassment by different offender types using various channels. Offender types were (1) individuals who repeatedly harassed victims (individual) and (2) the unspecified public (collective). The channels were public and private messages, e.g., public comments include comments on weblogs and replies on social media; private messages include direct messages on social media and e-mails. We asked about victimization through direct online harassment according to the following activities: my appearance has been abused (looks, body shape, etc.); my ability has been abused (knowledge, etc.); my personality has been abused; I received obscene messages (texts, images, etc.); I was threatened; and I have received requests for a sexual relationship. We call them "appearance abuse," "ability abuse," "personality abuse," "obscene messages," "intimidation," and "requesting a sexual relationship," respectively. The participants responded with "yes," "no," and "I do not remember/want to respond." We evaluated the number of "yes" answers in our analyses. In total, there were 24 questions (two offender types, two channels, and six victimization types).

Offenders can also harass victims indirectly (Sarna and Bhatia, 2017) through activities such as harmful gossiping about victims in other places, disclosing victims' personal information, sharing disinformation, and uploading photoshopped sexual images. We asked about victimization through indirect online harassment according to the following criteria: my photos and videos were

uploaded without authorization; my sexually suggestive photos and videos were uploaded, for example, zooming necklines and underclothes; my ungainly photographs and videos were uploaded, for example, rolling their eyes up into their heads; my photoshopped photos and videos were uploaded; my personal information was disclosed, for example, my residential address; I assisted in situations that were difficult for them to respond to, for example, suicidal feelings; my disinformation was spread on social media; and my disinformation was published on curation and news sites. Participants responded with "yes," "no," or "I do not remember/want to respond." We evaluated the number of "yes" answers in our analyses. There were eight questions.

Exploratory factor analysis (EFA) revealed five factors for these 32 items by using maximum likelihood estimation (MLE) and Promax rotation. We selected the number of factors using the Bayesian information criterion (BIC). The comparative fit index (CFI) (Bentler, 1990) was 0.800 and the root mean square error of approximation (RMSEA) (Steiger, 1990) was 0.11 (0.103, 0.117), where the square brackets indicate a 90% confidence interval. These factors were interpreted as harassment in public spaces, harassment in private spaces, intimidation, dating requests, and sexual harassment (Table 1).

2.2.2 Emotional injury

Online harassment injures victims' mental health (Bliuc et al., 2018). To measure such injury, we assessed emotional injury due to online harassment using the following prompts: I wanted to remove my account; I did not want to use the Internet; I felt less motivated to go out and interact socially; I had a sinking feeling for several days; I wanted to be truant or leave my job; I became suicidally depressed; I indulged in substance abuse; and I engaged in self-harm and wrecked some things. Participants responded based on a 4-point Likert-type scale, i.e., agree, slightly agree, slightly disagree, and disagree. There were eight questions.

Using EFA, we acquired one factor, i.e., emotional injury (Table 2), for these eight items by applying MLE and Promax rotation, from which we selected the number of factors using the BIC [CFI: 0.923, RMSEA: 0.128 (0.102, 0.156)].

2.2.3 Actions against online harassment

Victims can initiate several actions against online harassment, including talking to authorities and asking platforms for restrictions on and punishment of offender accounts (Every-Palmer et al., 2015). We investigated the actions taken against online harassment with the following prompts: I avoided and consigned weblog/social media posts to my staff; I restricted comments and replies on my weblog/social media; I blocked/muted harasser accounts on platforms; I reported harassers to platforms; I stopped updating my weblog/social media; I contacted an inquiry counter of the platforms; I talked family/friends/business friends; I reported my victimization to my talent agency; I went to mental health counselors; I discussed with my legal consultants; and I spoke to the police. Participants responded with "yes," "no," or "I do not remember/want to respond." We evaluated the number of "yes" answers in our analyses. There were 11 questions.

TABLE 1　Result of the EFA of direct and indirect online harassment.

Category	Item	F1	F3	F2	F4	F5
Direct (public-individual)	My personality has been abused	**0.81**	0.28	−0.10	−0.08	−0.20
Direct (public-collective)	My appearance has been abused	**0.79**	−0.10	0.08	0.24	−0.17
Direct (public-individual)	My appearance has been abused	**0.79**	−0.18	0.00	0.33	−0.12
Direct (public-collective)	My personality has been abused	**0.75**	0.31	0.07	−0.15	−0.24
Direct (public-individual)	My ability has been abused	**0.57**	0.24	−0.09	−0.03	0.08
Direct (public-collective)	My ability has been abused	**0.54**	0.43	−0.02	−0.14	0.04
Indirect	My disinformation was spread on social media	**0.49**	0.17	0.01	−0.04	0.08
Indirect	My photoshopped photos and videos were uploaded	**0.48**	−0.20	0.21	0.03	0.17
Direct (public-collective)	I received obscene messages	**0.43**	0.00	−0.07	0.38	−0.15
Indirect	My ungainly photographs and videos were uploaded	**0.40**	−0.01	0.16	0.28	-0.02
Indirect	My disinformation was published on curation and news sites	**0.39**	0.04	0.13	0.03	0.13
Indirect	My personal information was disclosed	**0.33**	−0.08	0.27	0.00	−0.02
Indirect	My sexually suggestive photos and videos were uploaded	**0.16**	−0.07	−0.13	0.06	0.09
Direct (private-collective)	My personality has been abused	0.06	**0.82**	0.04	0.17	−0.16
Direct (private-individual)	My personality has been abused	0.02	**0.81**	−0.04	0.29	−0.11
Direct (private-collective)	My ability has been abused	−0.05	**0.77**	0.00	0.16	0.14
Direct (private-individual)	My ability has been abused	−0.08	**0.74**	−0.06	0.27	0.19
Indirect	I assisted in situations that were difficult for them to respond to	0.16	**0.25**	−0.01	−0.25	0.12
Direct (private-collective)	I was threatened	−0.07	−0.01	**0.92**	0.02	0.13
Direct (public-collective)	I was threatened	0.02	0.05	**0.87**	0.03	−0.06
Direct (private-individual)	I was threatened	−0.05	−0.07	**0.86**	0.07	0.14
Direct (public-individual)	I was threatened	0.12	-0.01	**0.74**	−0.03	−0.04
Indirect	My photos and videos were uploaded without authorization	0.12	0.10	**0.16**	0.16	0.12
Direct (private-individual)	My appearance has been abused	0.01	0.18	0.06	**0.85**	−0.01
Direct (private-collective)	My appearance has been abused	−0.02	0.17	0.05	**0.73**	0.06
Direct (private-individual)	I received obscene messages	−0.03	0.09	0.02	**0.57**	0.09
Direct (private-collective)	I received obscene messages	0.07	0.05	0.02	**0.44**	−0.02
Direct (public-individual)	I received obscene messages	0.40	0.01	−0.14	**0.41**	−0.01
Direct (private-individual)	I have received requests for a sexual relationship	−0.20	0.04	−0.01	0.13	**0.84**
Direct (public-collective)	I have received requests for a sexual relationship	−0.16	0.08	0.02	0.07	**0.80**
Direct (public-individual)	I have received requests for a sexual relationship	0.08	−0.06	0.07	−0.13	**0.78**
Direct (public-collective)	I have received requests for a sexual relationship	0.03	0.02	0.05	−0.02	**0.72**
Factor correlations		F1	0.45	0.41	0.48	0.39
		F3		0.39	0.19	0.35
		F2			0.26	0.38
		F4				0.24

Bold values mean max values among each item's coefficients.

Through EFA, we acquired two factors for these 11 items by applying MLE and Promax rotation, from which the number of factors was selected using the BIC [CFI: 0.956, RMSEA: 0.043 (0.000, 0.070)]. These factors were understood to discuss victimization and the use of the functions provided by platforms (Table 3).

2.2.4 Offline harassment

Victims of online harassment are more likely to experience victimization in the real world (Every-Palmer et al., 2015; Meloy and Amman, 2016). We assessed offline victimization of stalking based on the following: I received persistent requests for dating and companionship; I was stalked, ambushed, and intruded on; I was

TABLE 2 Result of the EFA of emotional injury.

Item	F1
I did not want to use the Internet	0.83
I wanted to remove my account	0.83
I felt less motivated to go out and interact socially	0.81
I had a sinking feeling for several days	0.80
I wanted to be truant or leave my job	0.79
I became suicidally depressed	0.69
I indulged in substance abuse	0.52
I engaged in self-harm and wrecked some things	0.37

TABLE 3 Result of the EFA of actions against online harassment.

Item	F2	F1
I discussed with my legal consultants	0.70	−0.08
I spoke to the police	0.67	−0.12
I went to mental health counselors	0.49	−0.05
I contacted an inquiry counter of the platforms	0.42	0.09
I reported my victimization to my talent agency	0.41	−0.01
I talked family/friends/business friends	0.27	0.10
I avoided and consigned weblog/social media posts to my staff	0.22	0.02
I blocked/muted harasser accounts on platforms	−0.08	0.90
I restricted comments and replies on my weblog/social media	−0.15	0.65
I reported harassers to platforms	0.07	0.61
I stopped updating my weblog/social media	0.05	0.14
Factor correlations	F2	0.47

TABLE 4 Result of the EFA of offline harassment.

Item	F1
I was stalked, ambushed, and intruded on	0.87
I was followed by a stranger	0.79
I received persistent requests for dating and companionship	0.68
I received unwanted letters and/or presents	0.66
I was informed that I was being monitored	0.60
I uninterruptedly received calls, faxes, and e-mails	0.50

followed by a stranger; I was informed that I was being monitored; A GPS device was attached or tracking applications were installed without any prior consent; I uninterruptedly received calls, faxes, and e-mails; I received unwanted letters and/or presents; I was sent body fluid and/or excrement; and I was bugged, and/or I received a present with a bugging device. Participants responded with "yes," "no," or "I don't remember/want to respond." In our analyses, we evaluated the number of "yes" answers.

There were nine questions. Note that no respondents with "yes" on the fifth, eighth, and ninth items. Therefore, we analyzed the other six items.

Using EFA, we acquired one factor, offline stalking (Table 4), for these six items to which we applied MLE and Promax rotation and from which we selected the number of factors using the BIC [CFI: 0.962, RMSEA: 0.098 (0.057, 0.141)].

2.2.5 Demographic information

The participants provided their demographic information, including gender (female, male, and "I do not respond") and age in ten-year intervals (under 20[3]; 20s, 30s, 40s, 50s, and over 60s; and "I do not respond"). We used this coarse age-level grading to avoid identifying famous people about whom much information is publicly available. In the following regression analyses, "male" is a reference category for gender. The numbers of women, men, and no response were 172, 39, and 2, respectively. The numbers of participants in the age levels from the under 20s, 20s, 30s, 40s, 50s, and over 60s groups were 0, 2, 57, 91, 41, and 6, respectively. "I do not respond." was selected 16 times.

2.2.6 Number of online followers

The participants provided the total number of their online followers (weblogs, social media, video streaming services, etc.) in four levels (5,000 or less, 5,001–10,000, 10,001–100,000, and over 100,000). The numbers of these levels were 62, 36, 92, and 23, respectively.

2.2.7 Media appearance

The participants detailed how much exposure they experienced on conventional and online media e.g., TV, radio, newspapers, magazines, online video media, and online text media. This was based on 4-point Likert-type scales, i.e., high, moderate, low, and none.

2.3 Statistical model

To detail the associations between victims' features (demographic information, number of online followers, media exposure), direct and indirect victimization, offline stalking, emotional injury, and actions against harassment, we analyzed a structural equation model (SEM; Figure 1). We considered demographic information, the number of followers (fame and influence on social media), and mass/online media exposure (fame depending on media types) as victim features. The relationships between these victim features and online harassment victimization can explain who is vulnerable and how. The victimization should have damaged the victims emotionally and they did or did not act against the harassment. Therefore, we analyzed the associations between each victim's features, victimization, emotional injury, and actions against harassment. We used the EFA results for the variables of direct and indirect victimization, offline stalking, emotional injury, and actions against harassment. We excluded the participants who did not respond to gender and/or age from this

---

3 We instructed participants under 15 years to respond with their parents or guardians.

FIGURE 1
Overview of the SEM.
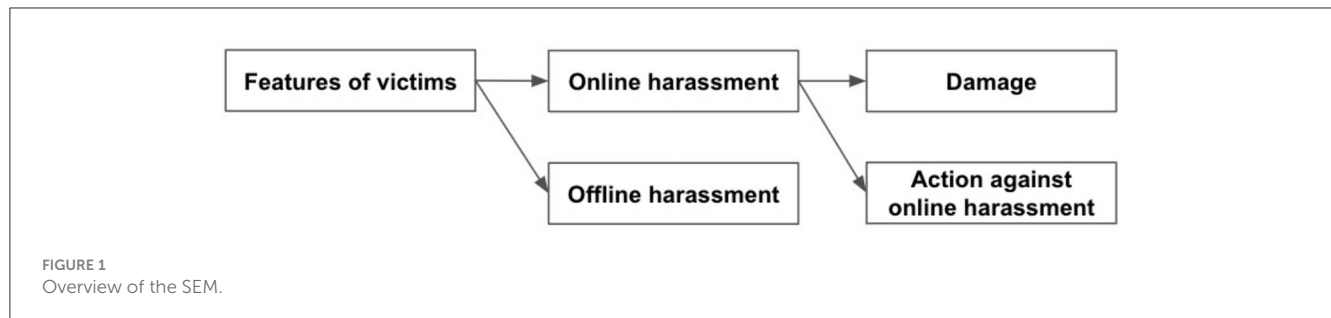
analysis. The number of participants in this analysis was 197. We used age levels as an ordinal measure with normalization.

This model was used to analyze the types of people vulnerable to particular types of online harassment, the types of online harassment harmful to victims' mental health, and the actions initiated by victims. Offline and online harassment were examined and compared.

2.4 Ethics statement

Our study was approved by the ethics committee of Graduate School of Sociology, Toyo University (P230067). Furthermore, all procedures were conducted in accordance with the guidelines for studies involving human participants and the ethical standards of the institutional research committee. Informed consent was obtained online from the participants for participation in our online survey, and the purpose of this study was fully explained (see Supplementary Figures S1, S2). The participants were over 18 years of age. The participants allowed the authors to analyze their data and anti-harassment actions for academic purposes on condition that the quantitative data outputs were aggregated, meaning no identifying information was presented.

CyberAgent, Inc. facilitated this questionnaire survey based on its privacy policy[4] and the informed consent for the survey. The survey was anonymized.

3 Results

3.1 Online harassment victimization

Figure 2 displays the frequencies of direct harassment according to offender types and channels (see Supplementary Table S1 for details). Offenders resorted to direct harassment in public forums (the victimization ratio was 43%), such as replying to social media and weblog comments. Most of these messages were abusive about physical or psychological characteristics. Victims tended to be targeted by both types of offenders, i.e., particular repeat offenders (individual) and unspecified masses (collective). Toxic messages were conveyed to victims also through private messages (the victimization ratio was 27%). Forty-nine percent of the participants reported being subjected to at least one type of direct harassment.

Figure 3 displays the frequencies of indirect harassment (see Supplementary Table S2 for details). The participants were exposed to indirect harassment, such as negative gossip, and disinformation (spread on social media and published on curation and news sites). Offenders shamed victims with unflattering, sexual, and photoshopped images and disclosed victims' personal information. Additionally, the victims were sometimes required to assist by answering questions that were difficult to respond to, e.g., suicidal feelings. Forty-nine percent of the participants reported being subjected to at least one type of indirect harassment.

Victims felt that online harassment negatively affected their mental health (Figure 4; see Supplementary Table S3 for details). The harassment resulted in them avoiding using the Internet (53%), caused "sinking feelings" (51%), led them to almost remove their accounts (33%), and instilled the desire to skip or leave their work (33%). The values in parentheses are "agree and slightly agree" answer ratios of the participants. Twenty-eight percent of victims complained of intense mental injury (suicidal feelings, self-harm reckless behavior, and intemperance/substance abuse.

Victims often used anti-harassment functions provided by platforms (Figure 5; see Supplementary Table S4 for details), such as blocking/muting offender accounts, closing comment forms, restricting accounts from posting comments, and reporting offender accounts to platforms to ban offenders.

Victims talked about their victimization to close people (family, friends, and business friends). They also spoke to their associated organizations, i.e., talent agencies and platforms.[5] A small number of victims consulted the police and legal counsel.

A few victims stopped their weblog/social media activities completely.

Online harassment victims were also stalked offline (Figure 6; see Supplementary Table S5 for details). Victims of online harassment were 3.1 times more likely to be stalked than those who were not.

3.2 Vulnerable people, emotional injury, and actions against online harassment

Figure 7 and Table 5 show the results of the SEM.

People with a modest number of followers are targeted for harassment in public spaces, sent dating requests, and sexually harassed. This phenomenon could be attributed to offenders not

---

4   https://www.cyberagent.co.jp/way/security/privacy/

5   Normally, several platforms seem to work with famous people.
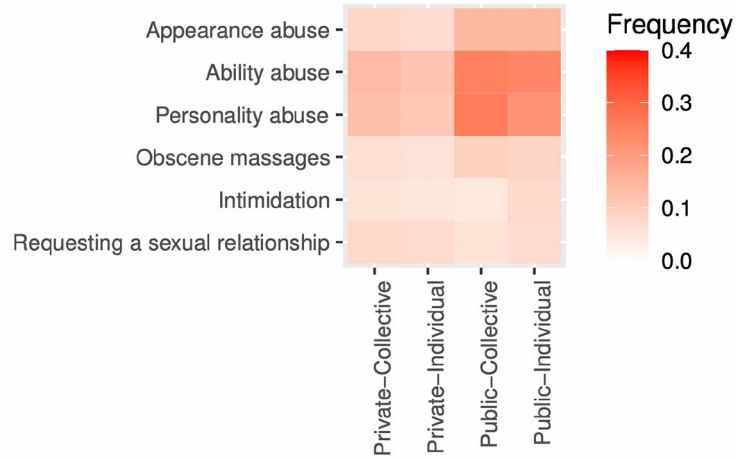
FIGURE 2
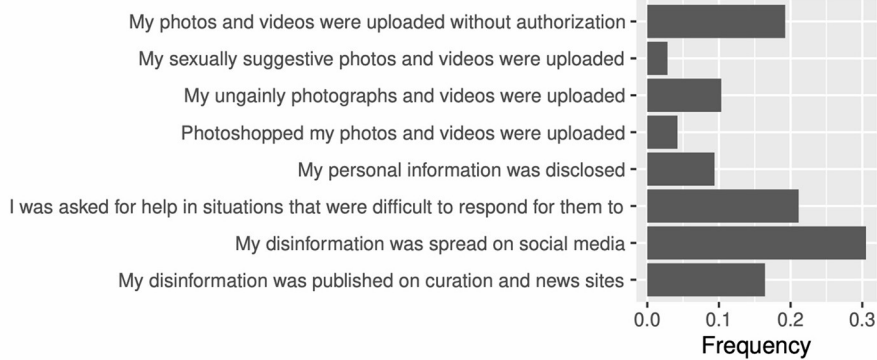Direct harassment frequencies of offender types and methods.

FIGURE 3
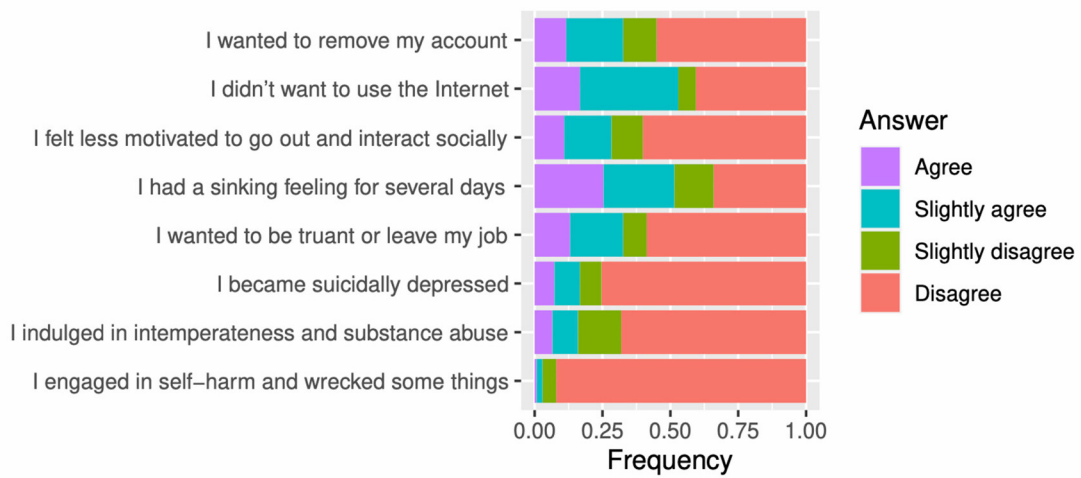Indirect harassment frequencies.
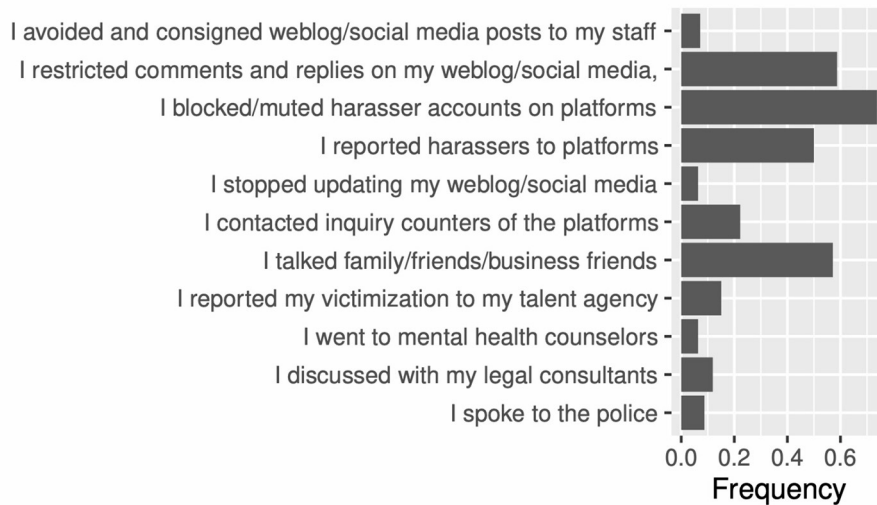

FIGURE 4
Emotional damages.
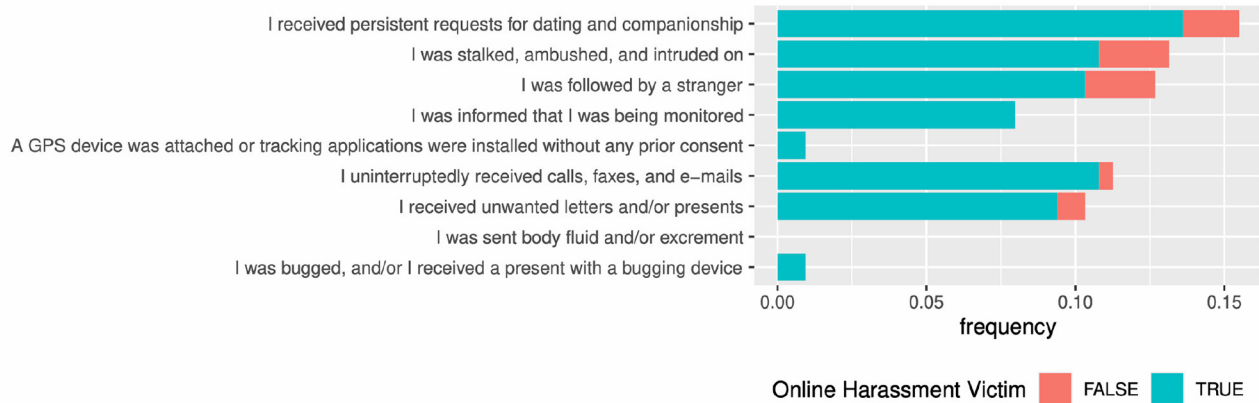
FIGURE 5
Action against online harassment.

FIGURE 6
Association between offline stalkings and online harassment.

finding opportunities to interact with people with a few followers and hesitating to attack people with many followers because such people wield tremendous influence on the Internet. Offline harassment showed similar tendencies to online harassment.

People with many mass media appearances were targets of all types of online harassment, excluding sexual harassment. By contrast, online media appearances did not reveal considerable effects, excluding harassment in public spaces.

Younger people may be targeted for harassment in private spaces and with dating requests. There were insignificant gender differences in online harassment. Although offline harassment followed online harassment trends, women were more vulnerable to offline stalking than men.
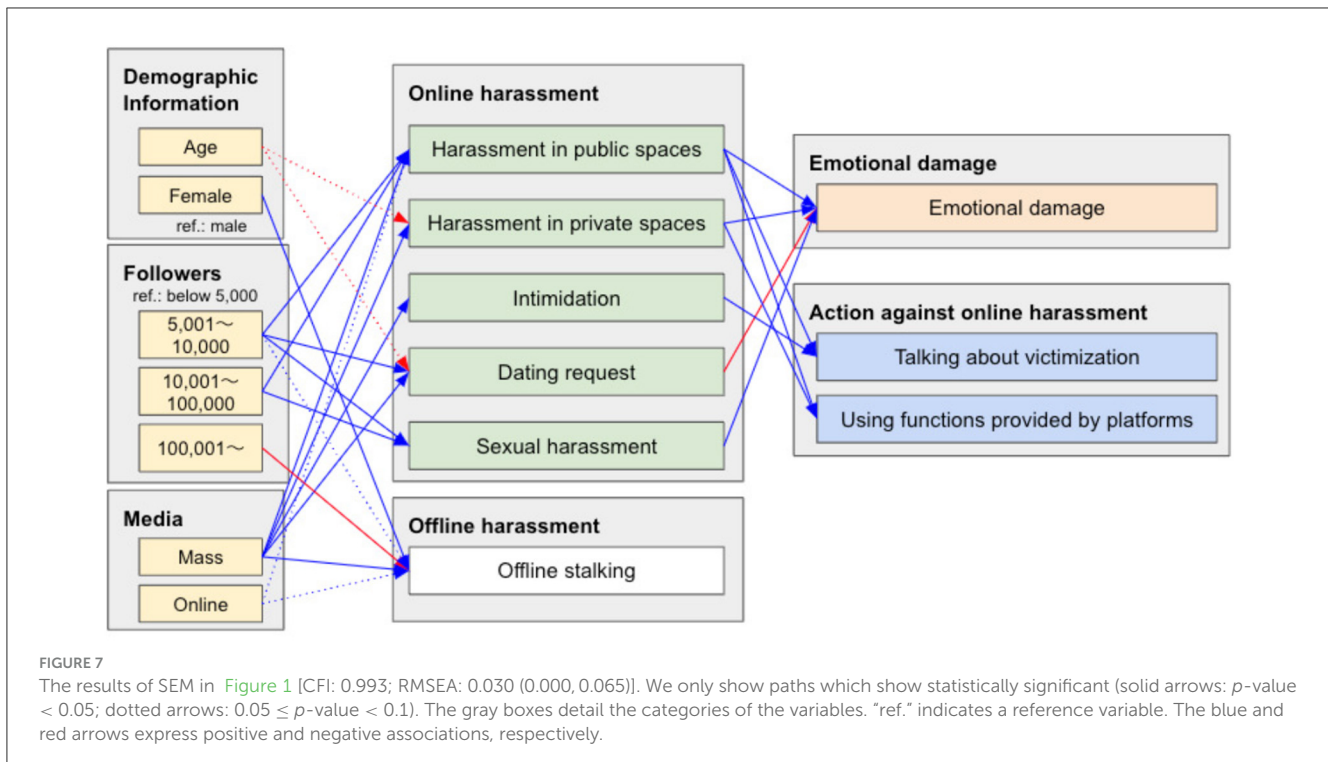
Victims of sexual harassment and harassment in public/private spaces were emotionally damaged. Reported experiences of undesirable dating requests did not cause the victims severe injury. This seems to be because persistent and vulgar dating requests could be considered sexual

harassment. Intimidation also did not exhibit associations with emotional injury.

Actions against online harassment change depending on the type of victimization.

Victims used anti-harassment functions provided by platforms (e.g., blocking/reporting accounts and closing comment forms) in public and private spaces. By contrast, intimidation, dating requests, and sexual harassment were insignificant in triggering the use of anti-harassment functions. This may be because anti-harassment functions do not remove the fear of intimidation; victims hesitate to block unwanted messages without explicit hostility, such as dating requests and sexual harassment.

Victims of harassment in public spaces and intimidation tended to raise this with others. They voice this harassment because anyone can observe harassment in public spaces, and intimidation is a crime. By contrast, harassment in private spaces and sexual harassment had insignificant effect on voicing about victimization, even if these activities were harmful.

FIGURE 7
The results of SEM in Figure 1 [CFI: 0.993; RMSEA: 0.030 (0.000, 0.065)]. We only show paths which show statistically significant (solid arrows: *p*-value < 0.05; dotted arrows: 0.05 ≤ *p*-value < 0.1). The gray boxes detail the categories of the variables. "ref." indicates a reference variable. The blue and red arrows express positive and negative associations, respectively.

To analyze the relationships between voicing victimization and whom they talk to, logistic regressions were conducted, in which the explained variables were whom they talked to; the explanatory variables were types of online harassment; and the control variables were age levels, and gender. Table 6 shows the results of these regressions. The explanatory and control variables were selected by Akaike's information criterion (AIC).

Victims talked about harassment in public spaces, regardless of whom they talked to. By contrast, in the case of talking about harassment in private spaces, they talked to were limited (platforms). Remarkably, victims did not seem to speak out about this harassment to the police. Regarding intimidation, victims contacted authorities with concrete legal actions (talent agencies, legal consultants, and police). Victims also seemed to talk about it to close people (family, friends, and business friends). Dating request victimization tended to be discussed with close people. However, no tendencies regarding sexual harassment revealed with whom it was discussed, i.e., it was not selected by AIC.

4 Discussion

4.1 Picture of online harassment victimization

Japanese famous people were harmed by various forms of online harassment, Victimization ratios of direct harassment in public and private places were 43 and 27%, respectively. That of indirect harassment was 49%. These victimization ratios are higher than Japanese journalists (21.5%) (Yamaguchi et al., 2023a) and ordinary people (4.7%) (Yamaguchi et al., 2023b). This suggests a famous risk in Japan even with considering participation bias of

our research. Online harassment victims tended to be harmed by offline stalking, as in the case of certain New Zealand politicians (Every-Palmer et al., 2015).

Most Japanese studies on online harassment are case studies, such as the analysis of individual-specific cases (Tonami et al., 2022) and flame war cases (Tanaka and Yamaguchi, 2021). The comprehensive picture of online harassment victimization is not clear. Recently, reports have been published on the surveys of victimization of ordinary people and journalists (Yamaguchi et al., 2023a,b). However, not much is known about celebrities who are more likely to be targeted. This study is the first to reveal the picture of victimization among the celebrity population. It can contribute to complementing the picture of online harassment victimization.

Victims were harmed by online harassment, particularly harassment in public/private spaces and sexual harassment. Some of the victims were severely mentally stressed and had negative emotions, such as suicidal thoughts, self-harm/wrecking, and substance abuse. However, the many of them did not stop their weblog/social media activities because it can be important for business (Van den Bulck et al., 2014). Thus, victims had to use online media even if they faced online harassment.

Frequent appearances in mass media and having more than a certain number of followers seemed to provide opportunities for offenders to target and harass victims. By contrast, having many followers seemed to inhibit being targeted for harassment because offenders seem to fear counterattacks from the victims' followers; for example, if victims criticize offenders' accounts and posts, victims' followers may also criticize offenders. Gender differences in online harassment victimization were not revealed. This contrasted with offline harassment, in which women were more targeted. This may be because offline offenders fear physical counterattacks; however, online offenders can harass victims without fear of

TABLE 5 Details of the SEM result.

Explained variable	Explanatory variable	Coef.	SE
Harassment in public spaces	Female	0.17	0.13
	Age	−0.01	0.06
	Followers: 5,001–10,000	0.50	0.20*
	Followers: 10,001–100,000	0.32	0.13*
	Followers: 100,001–	0.26	0.22
	Mass media	0.28	0.10**
	Online media	0.16	0.09·
Harassment in private spaces	Female	0.08	0.15
	Age	−0.11	0.06·
	Followers: 5,001-10,000	0.08	0.20
	Followers: 10,001–100,000	0.18	0.17
	Followers: 100,001–	-0.21	0.23
	Mass media	0.34	0.10***
	Online media	−0.01	0.09
Intimidation	Female	−0.27	0.21
	Age	0.04	0.06
	Followers: 5,001–10,000	0.39	0.24
	Followers: 10,001–100,000	0.14	0.13
	Followers: 100,001–	0.23	0.33
	Mass media	0.32	0.16*
	Online media	−0.10	0.12
Dating request	Female	−0.16	0.20
	Age	−0.11	0.06·
	Followers: 5,001–10,000	0.66	0.28*
	Followers: 10,001–100,000	0.02	0.11
	Followers: 100,001–	0.06	0.24
	Mass media	0.24	0.10**
	Online media	−0.05	0.08
Sexual harassment	Female	0.13	0.16
	Age	−0.02	0.06
	Followers: 5,001–10,000	0.50	0.21*
	Followers: 10,001–100,000	0.26	0.12*
	Followers: 100,001–	0.38	0.32
	Mass media	0.14	0.11
	Online media	0.04	0.09
Offline stalking	Female	0.42	0.13***
	Age	0.02	0.06
	Followers: 5,001–10,000	0.43	0.23·
	Followers: 10,001–100,000	−0.18	0.13
	Followers: 100,001–	-0.40	0.19*
	Mass media	0.21	0.10*

(Continued)

TABLE 5 (Continued)

Explained variable	Explanatory variable	Coef.	SE
	Online media	0.18	0.10·
Emotional injury	Harassment in public spaces	0.27	0.10**
	Harassment in private spaces	0.20	0.08*
	Intimidation	0.06	0.07
	Dating request	−0.16	0.07*
	Sexual harassment	0.17	0.08*
Talking about victimization	Harassment in public spaces	0.49	0.12***
	Harassment in private spaces	−0.05	0.06
	Intimidation	0.26	0.08***
	Dating request	0.08	0.08
	Sexual harassment	−0.04	0.08
Using functions provided by platforms	Harassment in public spaces	0.34	0.07***
	Harassment in private spaces	0.22	0.05***
	Intimidation	−0.06	0.06
	Dating request	0.05	0.06
	Sexual harassment	0.00	0.05

***, **, *, and · indicate significant differences at p−values $\leq 0.001, 0.01, 0.05$, and 0.1.

physical harm (Chapman, 1995). Thus, offenders may choose targets, without the fear of counterattacks. This is consistent with the fact that journalists who do not have the protection of an organization are more likely to be targeted (Yamaguchi et al., 2023a). Note that the gender difference might have appeared because female participants were more than male participants, i.e., there may be participation bias.

Victims' tendencies of discussing victimization and using anti-harassment functions provided by weblogs and social media were investigated. A previous study in Japan (Yamaguchi et al., 2023a) pointed out that the number of people who take concrete action when victimized is limited. We suggest why it is challenging to take concrete actions against online harassment as follows. Victims appeared hesitant to speak out about harassment through private methods, such as direct messages, compared with comments in public spaces because victims must disclose the abuse against them. Victims were not required to complain about comments in public spaces because they were inherently observable. Victims are hesitant to disclose dating requests and sexual harassment because such sexual victimization aggravates victims' stigma anxieties and their fear of public rejection (Mesch and Talmud, 2006; Andalibi et al., 2016; Foster and Fullagar, 2018).

In the following, we discuss the consistencies and differences with previous studies. Chen et al. (2018) and Yamaguchi et al. (2023a) indicated the importance of potential victims having countermeasures, e.g., training to handle harassment and the organizations protecting them from abuse. In addition to them, our findings suggest that noticing such countermeasures may inhibit offensive behavior. Although studies of famous female victims (Chen et al., 2018; Tonami et al., 2022) show that sexism is one of the major factors in online harassment, there is no significant

TABLE 6 Relationships between talking about victimization and its partners.

Explanatory variable	Platform		Family etc.		Talent agency		Counselor		Legal consultant		Police	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Intercept	−2.08	0.25***	−1.95	0.53***	−3.67	0.85***	−4.28	0.64***	−3.40	0.47***	−3.73	0.52***
Harassment in public spaces	0.93	0.22***	0.49	0.20*	0.83	0.23**	0.86	0.37*	1.10	0.27***	1.10	0.30***
Harassment in private spaces	0.39	0.19*									−0.78	0.48·
Intimidation			0.50	0.26·	0.35	0.20·	0.42	0.22	0.46	0.19*	0.62	0.28*
Dating request			0.49	0.24*								
Sexual harassment												
Age					0.41							
Gender: female			1.95	0.58**	1.16	0.87						
McFadden's R^2	0.20		0.16		0.23		0.28		0.36		0.31	

***, **, *, and · indicate significant differences at $p-$values $\leq$ 0.001, 0.01, 0.05, and 0.1. Blank cells represent variables not selected by AIC.

difference in terms of gender in online harassment. This suggests that certain vulnerabilities are gender-specific to females and males.

4.2 Issues of online harassment

We identified six issues in online harassment.

First, with respect to dating requests, sexual harassment, and intimidation, victims do not tend to use anti-harassment functions provided by platforms, such as blocking/reporting offender accounts and closing comment forms.

Second, victims hesitate to discuss dating requests, sexual harassment, or harassment through private messaging.

Third, offenders harm victims indirectly through activities such as sordid gossip, disclosing and fabricating personal information, and sexualized photoshopped images. The number of victims of indirect harassment was close to that of direct harassment. Victims and platforms cannot remove these posts easily because the removal of such content requires complex procedures.

Fourth, people who seem unwilling to fight back tend to be targeted by offenders. Offenders use the characteristic features of online communication, which remove the fear of physical harm and online backlash.

Fifth, victims receive stressful messages without the sender's malicious intent, for example, sensitive self-disclosure of suicidal feelings. Removing such comments and banning such accounts by platforms is difficult because this approach usually does not infringe on the terms of service.

Sixth, online harassment causes victims considerable harm.

4.3 Approaches against six issues

We can consider several potential approaches to overcome these concerns.

Facilitating re-consideration of posting comments is effective in discouraging offensive posts (Katsaros et al., 2022). This approach

has been adopted by many platforms on public comment forums.[6] Our survey revealed that victims receive toxic private messages that harm them. Victims hesitate to speak about such victimizations. Thus, adopting the re-consideration function for private message forms is effective.

Changing re-consideration policies to offenders depending on the types of online harassment can reduce some types of online harassment. For example, the incidences of intimidation, in which victims do not use anti-harassment functions provided by platforms, may decrease by indicating that intimidation is a criminal act. Flagging that comments and messages may include sexual harassment could discourage sexual harassment (Mcdonald et al., 2014; Blackwell et al., 2017).

A step-by-step guide including the definitions of online harassment, and concrete examples should be issued against online harassment to support victims' actions against online harassment and mitigate online harassment. Such a guide can facilitate the understanding of complex legal procedures for stopping toxic messages. This guide should clarify the scope of online harassment by defining it and providing concrete examples. This clarification can reduce victims' hesitancy to initiate actions against online harassers when they understand that this guide defines the received toxic and sexual messages as harassment (Blackwell et al., 2017). The guide can affect offenders' behavior. Offenders assume that attacking celebrities will not have any repercussions because celebrities are socially invulnerable (Lawson, 2017; Ouvrein et al., 2018; Lee et al., 2020). Online harassment may decrease if victims become aware that their behavior constitutes harassment

---

6 For example, https://about.fb.com/news/2019/12/our-progress-on-leading-the-fight-against-online-bullying/, https://blog.nextdoor.com/2019/09/18/announcing-our-new-feature-to-promote-kindness-in-neighborhoods/, https://blog.youtube/news-and-events/make-youtube-more-inclusive-platform/, https://medium.com/jigsaw/helping-authors-understand-toxicity-one-comment-at-a-time-f8b43824cf41, https://newsroom.pinterest.com/en/creatorcode, https://newsroom.tiktok.com/en-us/new-tools-to-promote-kindness, and https://ameblo.jp/staff/entry-12612189833.html.

(Blackwell et al., 2017). Furthermore, cautioning offenders about the risk of online harassment (e.g., banning and litigation) can discourage toxic behavior. For example, although the anonymity of the Internet facilitates offensive behavior (Santana, 2014; Bliuc et al., 2018; Takano et al., 2021), this guide informs offenders that they can be exposed through legal procedures. Thus, vulnerable people are assured of countermeasures against offenders. As a supplemental approach, online legal consultation, i.e., support from specialists, can be helpful. Displaying attitudes adopted by platforms to reject online harassment also inhibits toxic posts (Kang et al., 2022).

The guide should share the definition and adverse effects of online harassment with people (Blackwell et al., 2017). This could mitigate the stigma of online harassment (Mesch and Talmud, 2006; Andalibi et al., 2016; Foster and Fullagar, 2018), victim-blaming (Scott et al., 2019; Hand, 2021), and propagating offensive behavior (Bliuc et al., 2018; Yokotani and Takano, 2021). Educating people accordingly can ameliorate the adverse effects of sensational coverage (Kang et al., 2022).

Insurances against online harassment can encourage victims to speak out against it. The insurances provide financial support and lawyer referral services. This may discourage toxic behavior because of increased litigation risks.

For sensitive messages, for example, suicidal feelings, platforms can place a report button on their portal. The platform should introduce inquiry counters to senders of these messages, such as mental health counseling. This measure could be the simplest initial support for the senders of such messages. This measure could be expected to decrease the victims' stress because it serves the sender's interests. Such mental support for senders (offenders) can decrease offensive behavior because they tend to have a psychiatric disorder (Meloy and Amman, 2016).

Online counseling is effective for victims' mental health in mitigating the injury of online harassment (Kraus et al., 2010).

These measures have been partially implemented by a Japanese weblog platform, Ameba.[7] For example, they defined online harassment and the penalties levied by the platform against such harassment. This also prepared a guide for actions against online harassment, provided/introduced inquiry counters, and facilitated re-consideration in private messages in addition to public comment forms. Other options are also considered.

These suggestions need to be verified through field experiments on actual platforms to confirm if they can effectively resolve the issues identified in this study.

4.4 Limitations and future works

Participants consisting of famous people were recruited by a Japanese Internet company. The number of participants was not sufficiently large. Additionally, there is a possibility of bias due to the lack of comprehensiveness in the survey items on direct and indirect harassment. More representative sampling and more exhaustive items of questions would generalize the findings of this study and provide greater insight.

Our findings could be applied to others and not just Japanese famous people. Examining this applicability should contribute to broader online harassment prevention.

The characteristics of online harassment victimization are common across many cultures and countries (Chen et al., 2018). Examining our results in various countries and regions could provide more extensive knowledge.

The effectiveness of the proposed approaches on victims, offenders, and society should be examined.

5 Summary

We investigated online harassment victimization among famous people and their responses to such victimization. Several approaches were proposed for inhibiting harassment and mitigating victims' burdens. The victimization of famous people can negatively affect their social life. Online spaces, such as the metaverse and live streaming, have numerous highly detailed interactions, which may increase the risk of online harassment (Wiederhold, 2022). Our research can contribute to platforms establishing support systems against online harassment from various perspectives, such as supporting victims, inhibiting toxic expressions, and educating people.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Graduate School of Sociology, Toyo University Research Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MT: Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. FT: Writing – review & editing. CO: Writing – review & editing, Conceptualization. NN: Writing – review & editing.

Funding

Conflict of interest

MT is an employee of CyberAgent, Inc.

FT, CO, and NN were funded by CyberAgent, Inc. The funder had the following involvement in the study: data collection

---

7 https://ameblo.jp/staff/entry-12732411709.html

(we requested cooperation from potential respondents through CyberAgent, Inc.). This study was conducted as a part of the work of Cyberagent, Inc., an Internet service provider, and its collaborator Chiki Lab, motivated to improve the online environment. There are no patents, products in development, or marketed products to declare.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1386146/full#supplementary-material

References

Andalibi, N., Haimson, O. L., De Choudhury, M., and Forte, A. (2016). "Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI '16* (New York, NY: ACM Press), 3906–3918. doi: 10.1145/2858036.2858096

Badjatiya, P., Gupta, M., and Varma, V. (2019). "Stereotypical bias removal for hate speech detection task using knowledge-based generalizations," in *The Web Conference 2019 – Proceedings of the World Wide Web Conference, WWW 2019* (New York, NY: Association for Computing Machinery, Inc), 49–59. doi: 10.1145/3308558.3313504

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

BIGLOBE Inc (2020). *80% of People Think That Penalties for Slander on SNS Should be Strengthened*. Available online at: https://www.biglobe.co.jp/pressroom/info/2020/08/200826-1 (accessed February 26, 2024).

Blackwell, L., Dimond, J., Schoenebeck, S., and Lampe, C. (2017). "Classification and its consequences for online harassment," in *Proceedings of the ACM on Human-Computer Interaction* (New York, NY: ACM), 1. doi: 10.1145/3134659

Bliuc, A.-M., Faulkner, N., Jakubowicz, A., and McGarty, C. (2018). Online networks of racial hate: a systematic review of 10 years of research on cyber-racism. *Comput. Human Behav.* 87, 75–86. doi: 10.1016/j.chb.2018.05.026

Brailovskaia, J., Teismann, T., and Margraf, J. (2018). Cyberbullying, positive mental health and suicide ideation/behavior. *Psychiatry Res.* 267, 240–242. doi: 10.1016/j.psychres.2018.05.074

Chapman, G. (1995). Flamers. *New Republic* 212, 13–15.

Chen, G. M., Pain, P., Chen, V. Y., Mekelburg, M., Springer, N., Troger, F., et al. (2018). 'You really have to have a thick skin': a cross-cultural perspective on how online harassment influences female journalists. *Journalism* 21, 877–895. doi: 10.1177/1464884918768500

Cho, K. (2017). Quantitative text analysis of "Yahoo! News": focusing on comments on Koreans. *J. Appl. Sociol.* 59, 113–127.

Crandall, C. S., and Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychol. Bull.* 129, 414–446. doi: 10.1037/0033-2909.129.3.414

Dooley, B., and Hida, H. (2020). After Reality Star's Death, Japan Vows to Rip the Mask Off Online Hate. *The New York Times*. Available online at: https://www.nytimes.com/2020/06/01/business/hana-kimura-terrace-house.html (accessed February 26, 2024).

Dooley, B., and Ueno, H. (2022). "In the U.S., His Site Has Been Linked to Massacres," in *Japan, He's a Star*. The New York Times. Available online at: https://www.nytimes.com/2022/12/18/business/4chan-hiroyuki-nishimura.html (accessed February 26, 2024).

Every-Palmer, S., Barry-Walsh, J., and Pathé, M. (2015). Harassment, stalking, threats and attacks targeting new zealand politicians: a mental health issue. *Aust. N. Z. J. Psychiatry* 49, 634–641. doi: 10.1177/0004867415583700

Foster, P. J., and Fullagar, C. J. (2018). Why don't we report sexual harassment? An Application of the theory of planned behavior. *Basic Appl. Soc. Psych.* 40, 148–160. doi: 10.1080/01973533.2018.1449747

Hand, C. (2021). Why is celebrity abuse on twitter so bad? It might be a problem with our empathye. *Conversation* 3. Available online at: https://theconversation.com/why-is-celebrity-abuse-on-twitter-so-bad-it-might-be-a-problem-with-our-empathy-154970

Hinduja, S., and Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Arch. Suicide Res.* 14, 206–221. doi: 10.1080/13811118.2010.494133

Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., et al. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* 573, 261–265. doi: 10.1038/s41586-019-1494-7

Kang, N. G., Kuo, T., and Grossklags, J. (2022). Closing Pandora's box on naver: toward ending cyber harassment. *Proc. Int. AAAI Conf. Web Soc. Media* 16, 465–476. doi: 10.1609/icwsm.v16i1.19307

Katsaros, M., Yang, K., and Fratamico, L. (2022). Reconsidering tweets: intervening during tweet creation decreases offensive content. *Proc. Int. AAAI Conf. Web Soc. Media* 16, 477–487. doi: 10.1609/icwsm.v16i1.19308

Kim, J. H., Park, E. C., Nam, J. M., Park, S., Cho, J., Kim, S. J., et al. (2013). The werther effect of two celebrity suicides: an entertainer and a politician. *PLoS ONE* 8:e84876. doi: 10.1371/journal.pone.0084876

Kraus, R., Stricker, G., and Speyer, C. (2010). *Online Counseling: A Handbook for Mental Health Professionals*, 2nd edition. Cambridge, MA: Academic Press.

Lawson, C. E. (2017). Innocent Victims, creepy boys: discursive framings of sexuality in online news coverage of the celebrity nude photo hack. *Fem. Media Stud.* 18, 825–841. doi: 10.1080/14680777.2017.1350197

Lawson, C. E. (2018). Platform vulnerabilities: harassment and misogynoir in the digital attack on leslie jones. *Inf. Commun. Soc.* 21, 818–833. doi: 10.1080/1369118X.2018.1437203

Lee, W., Lee, H., and Student, M. S. (2020). Bias & hate speech detection using deep learning: multi-channel CNN modeling with attention. *J. Korea Inst. Inf. Commun. Eng.* 24, 1595–1603. doi: 10.6109/jkiice.2020.24.12.1595

Marwick, A. E. (2017). Scandal or sex crime? Gendered privacy and the celebrity nude photo leaks. *Ethics Inf. Technol.* 19, 177–191. doi: 10.1007/s10676-017-9431-7

Matamoros-Fernández, A. (2017). Platformed racism: the mediation and circulation of an australian race-based controversy on Twitter, Facebook and YouTube. *Inf. Commun. Soc.* 20, 930–946. doi: 10.1080/1369118X.2017.1293130

Mcdonald, P., Charlesworth, S., and Graham, T. (2014). Developing a framework of effective prevention and response strategies in workplace sexual harassment. *Asia Pac. J. Hum. Resour.* 53, 41–58. doi: 10.1111/1744-7941.12046

Meloy, J. R., and Amman, M. (2016). Public figure attacks in the United States, 1995–2015. *Behav. Sci. Law* 34, 622–644. doi: 10.1002/bsl.2253

Mesch, G., and Talmud, I. (2006). Online friendship formation, communication channels, and social closeness. *Int. J. Internet Sci.* 1, 29–44.

Milosevic, T., Van Royen, K., and Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: new directions in the midst of complexity. *Int. J. Bullying Prev.* 4, 1–5. doi: 10.1007/s42380-022-00117-x

Ouvrein, G., De Backer, C. J., and Vandebosch, H. (2018). Joining the clash or refusing to bash? bystanders reactions to online celebrity bashing. *Cyberpsychology* 12:1. doi: 10.5817/CP2018-4-5

Park, S., and Kim, J. (2021). Tweeting about abusive comments and misogyny in south korea following the suicide of sulli, a female K-pop star: social and semantic network analyses. *Prof. Inf.* 30, e300505. doi: 10.3145/epi.2021.sep.05

Phillips, D. P. (1974). The Influence of suggestion on suicide: substantive and theroretical implications of the werther effect. *Am. Sociol. Rev.* 39, 340–354. doi: 10.2307/2094294

Rosen, G., Kreiner, H., and Levi-Belz, Y. (2019). Public response to suicide news reports as reflected in computerized text analysis of online reader comments. *Arch. Suicide Res.* 24, 243–259. doi: 10.1080/13811118.2018.1563578

Saleem, H. M., Dillon, K. P., Benesch, S., and Ruths, D. (2016). "A web of hate: tackling hateful speech in online social spaces," in *The Proceedings of Text Analytics for Cybersecurity and Online Safety* [Paris: European Language Resources Association (ELRA)].

Santana, A. D. (2014). Virtuous or vitriolic: the effect of anonymity on civility in online newspaper reader comment boards. *Journal. Pract.* 8, 18–33. doi: 10.1080/17512786.2013.813194

Sap, M., Card, D., Gabriel, S., Choi, Y., and Smith, N. A. (2019). "The risk of racial bias in hate speech detection," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Stroudsburg, PA: Association for Computational Linguistics), 1668–1678. doi: 10.18653/v1/P19-1163

Sarna, G., and Bhatia, M. P. (2017). Content based approach to find the credibility of user in social networks: an application of cyberbullying. *Int. J. Mach. Learn. Cybern.* 8, 677–689. doi: 10.1007/s13042-015-0463-1

Scott, G. G., Wiencierz, S., and Hand, C. J. (2019). The volume and source of cyberabuse influences victim blame and perceptions of attractiveness. *Comput. Human Behav.* 92, 119–127. doi: 10.1016/j.chb.2018.10.037

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav. Res.* 25, 173–180. doi: 10.1207/s15327906mbr2502_4

Takano, M., Taka, F., Morishita, S., Nishi, T., and Ogawa, Y. (2021). Three clusters of content-audience associations in expression of racial prejudice while consuming online television news. *PLoS ONE* 16:e0255101. doi: 10.1371/journal.pone.0255101

Takano, M., Taka, F., Morishita, S., Nishi, T., and Ogawa, Y. (2021). *The Research on Flaming*. Tokyo: Keisoshobou.

Tonami, A., Yoshida, M., and Sano, Y. (2022). Online harassment in Japan: dissecting the targeting of a female journalist. *F1000Res.* 10:1164. doi: 10.12688/f1000research.74657.2

Van den Bulck, H., Claessens, N., and Bels, A. (2014). 'By working she means tweeting': online celebrity gossip media and audience readings of celebrity twitter behaviour. *Celebr. Stud.* 5, 514–517. doi: 10.1080/19392397.2014.980655

Wiederhold, B. K. (2022). Sexual harassment in the metaverse. *Cyberpsychol. Behav. Soc. Netw.* 25, 479–480. doi: 10.1089/cyber.2022.29253.editorial

Yamaguchi, S., Ohshima, H., and Watanabe, T. (2023a). The Reality of Slander Against Japanese Journalists. *Innovation Nippon*. Available online at: http://www.innovation-nippon.jp/reports/2022IN_report_journalist_full.pdf (accessed February 26, 2024).

Yamaguchi, S., Tanihara, T., and Ohshima, H. (2023b). *Survey on the Reality of Slander in Japan*. Innovation Nippon. Available online at: : http://www.innovation-nippon.jp/reports/2022IN_report_hibou_full.pdf (accessed February 26, 2024).

Yokotani, K., and Takano, M. (2021). Social contagion of cyberbullying via online perpetrator and victim networks. *Comput. Human Behav.* 119:106719. doi: 10.1016/j.chb.2021.106719