frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 June 2024
DOI 10.3389/fpsyg.2024.1395668

OPEN ACCESS

EDITED BY
Michelle F. Wright,
Indiana State University, United States

REVIEWED BY
Junxiang Chen,
Indiana University, United States
Saerom Lee,
Kyungpook National University,
Republic of Korea
Mengru Sun,
Zhejiang University, China

*CORRESPONDENCE
Lifang Li
✉ lilf69@mail.sysu.edu.cn

RECEIVED 05 March 2024
ACCEPTED 15 May 2024
PUBLISHED 13 June 2024

CITATION
Li L, Zhou J, McManus S, Stewart R and Roberts A (2024) Social media users' attitudes toward cyberbullying during the COVID-19 pandemic: associations with gender and verification status.
Front. Psychol. 15:1395668.
doi: 10.3389/fpsyg.2024.1395668

Social media users' attitudes toward cyberbullying during the COVID-19 pandemic: associations with gender and verification status

Lifang Li[1]*, Jiandong Zhou[2], Sally McManus[3], Robert Stewart[4,5] and Angus Roberts[6]

[1]School of Journalism and Communication, Sun Yat-sen University, Guangzhou, China, [2]Division of Health Science, Warwick Medical School, University of Warwick, Coventry, United Kingdom, [3]Violence and Society Centre, City, University of London, Northampton Square, London, United Kingdom, [4]Department of Psychological Medicine, King's College London, London, United Kingdom, [5]South London and Maudsley NHS Foundation Trust, London, United Kingdom, [6]Department of Biostatistics and Health Informatics, King's College London, London, United Kingdom

Introduction: Social media platforms such as Twitter and Weibo facilitate both positive and negative communication, including cyberbullying. Empirical evidence has revealed that cyberbullying increases when public crises occur, that such behavior is gendered, and that social media user account verification may deter it. However, the association of gender and verification status with cyberbullying is underexplored. This study aims to address this gap by examining how Weibo users' gender, verification status, and expression of affect and anger in posts influence cyberbullying attitudes. Specifically, it investigates how these factors differ between posts pro- and anti-cyberbullying of COVID-19 cases during the pandemic.

Methods: This study utilized social role theory, the Barlett and Gentile Cyberbullying Model, and general strain theory as theoretical frameworks. We applied text classification techniques to identify pro-cyberbullying and anti-cyberbullying posts on Weibo. Subsequently, we used a standardized mean difference method to compare the emotional content of these posts. Our analysis focused on the prevalence of affective and anger-related expressions, particularly examining variations across gender and verification status of the users.

Results: Our text classification identified distinct pro-cyberbullying and anti-cyberbullying posts. The standardized mean difference analysis revealed that pro-cyberbullying posts contained significantly more emotional content compared to anti-cyberbullying posts. Further, within the pro-cyberbullying category, posts by verified female users exhibited a higher frequency of anger-related words than those by other users.

Discussion: The findings from this study can enhance researchers' algorithms for identifying cyberbullying attitudes, refine the characterization of cyberbullying behavior using real-world social media data through the integration of the mentioned theories, and help government bodies improve their cyberbullying monitoring especially in the context of public health crises.

KEYWORDS

cyberbullying, COVID, gender, verification status, emotional responses

1 Introduction

Social media enables positive and negative communication, including cyberbullying (Jatmiko et al., 2020). Cyberbullying is defined as the use of the internet to send harassing or threatening messages, or post humiliating comments (Hinduja and Patchin, 2010). The effects of cyberbullying are potentially more severe than those of physical or verbal bullying because wider audiences can be reached and materials may be accessed repeatedly (Li and Peng, 2022), resulting in victims potentially reliving denigrating experiences (Hinduja and Patchin, 2010). As online access grows, the number of people exposed to cyberbullying may also increase (Beran and Qing, 2005).

Cyberbullying perpetration is associated with problematic internet use, defined as the psychological, social, school or work difficulties experienced because of using the internet (Yudes et al., 2021). During the COVID-19 pandemic, quarantine led many people to rely more heavily on text messages and social media, with an increased risk of cyberbullying (Cheng et al., 2020; Babvey et al., 2021; Morales-Arjona et al., 2022; Yang et al., 2022). In China and Malaysia, patients with COVID-19, especially those 'super-spreaders' who were confirmed as positive in certain cities and exposed by the media, were aggressively cyberbullied (Patel, 2021; Ting and Shamsul, 2022). Victims of cyberbullying often experience mental health harms as a result, including depression (Yudes et al., 2021) and suicide (Hinduja and Patchin, 2010; John et al., 2018). Researchers suggest that it is necessary to research cyberbullying behavior extensively (Bansal et al., 2023).

Investigations into the factors associated with social media users' cyberbullying behavior could therefore benefit research and policy-making in this area: for example, efficient invention policy design and implementation can help reduce and mitigate the negative impact of cyberbullying. (1) Self-control theory has been utilized to argue potential gender differences in engaging in cyberbullying behavior, suggesting that females are less inclined to participate in cyberbullying compared to males (Griezel et al., 2012; Wong et al., 2018; Marr and Duell, 2021); (2) anonymity show positive influence in engaging cyberbullying behavior based on Barlett and Gentile Cyberbullying Model (BGCM) (Barlett et al., 2021a); and (3) individuals with low affective empathy demonstrated higher cyberbullying scores (Ang and Goh, 2010). More specifically, following General Strain Theory (GST), anger was found to be an important mediating factor of cybervictimization and cyberbullying (Ak et al., 2015).

However, few studies have integrated these factors (gender, anonymity, and affective) to find out their effects on cyberbullying attitudes using real-world social media data. Analyzing the interplay between gender, anonymity, and emotional factors is key to gain a comprehensive understanding of cyberbullying behaviors. There also lacks studies employing actual social media data for this purpose.

We aimed to fill these research gaps using the cyberbullying experiences of the COVID-19 patents during the COVID-19 pandemic, as previous study revealed a positive correlation between proximal experiences with COVID-19 and cyberbullying (Barlett et al., 2021b). Specifically, we would like to test social media users' attitudes of pro-and anti-cyberbullying towards the patients. This interest arises from the observation that COVID-19 patients, particularly those whose travel routes have been disclosed, have been targets of cyberbullying (Lian et al., 2022). Using social media (Weibo) data created during the COVID-19 pandemic, this study aimed to:

1. Investigate the relationship between gender and the likelihood of sharing pro-cyberbullying content on Weibo.
2. Explore the connections between the verification status, gender, and cyberbullying attitudes of Weibo users.
3. Analyze how verification status, gender, and the use of emotional words are associated with the number of reposts (retweets) for pro-cyberbullying and anti-cyberbullying content.

To achieve the research objectives mentioned above, this paper introduced relevant theories in Section 2, discussed research methods in Section 3, presented research results in Section 4, summarized research findings, corresponding theoretical and practical implications, and limitations in Section 5, and finally concluded the paper in the last section.

2 Theoretical background

2.1 Gender and cyberbullying

Self-control theory refers to the capacity to delay immediate gratification, manage negative emotions, sustain perseverance in fulfilling obligations, and restrain impulsive behaviors (Kochanska et al., 1996). It entails regulating emotions, beliefs, and actions to cultivate healthy interpersonal relationships. According to this theory, self-control differs based on individuals' gender, with males typically exhibiting lower levels (Wu et al., 2023). Historically, females have been socialized to conform to societal norms, promoting self-regulation and risk aversion, thereby reducing the likelihood of engaging in criminal behaviors (Wong et al., 2018). Self-control theory was applied to explain cyberbullying behavior (Stults and You, 2022). Consequently, researchers aimed to investigate how gender influences cyberbullying (Wang et al., 2019; Marr and Duell, 2021). However, findings in this regard are inconsistent.

One survey-based study found no significant difference of gender in cyberbullying behavior, with respect to COVID-19 experiences (Barlett et al., 2021c). Another group of studies have found males to be more likely to engage in cyberbullying (Barlett et al., 2021c), and other studies have found women to be more likely (Görzig and Ólafsson, 2013). Gender differences in cyberbullying victimization and suicidal ideation has also been found, with the association being stronger in women (Machimbarrena et al., 2018). However, there has been limited understanding of how gender influences the attitudes of males and females towards cyberbullying, despite attitudes being a crucial factor in predicting future cyberbullying behavior according to BCGM (Barlett, 2017). In this way, this study aimed to investigate whether females are less inclined than males to share pro-cyberbullying posts.

2.2 BGCM and social media users' verification status

The BGCM posits that perceptions of anonymity and the belief in the irrelevance of physical stature are two interconnected cognitive structures that forecast cyberbullying attitudes (Barlett and Gentile, 2012). It also suggests that having positive attitudes toward cyberbullying predicts future engagement in cyberbullying behaviors (Barlett et al., 2017).

In social media platform, verified users, i.e., those who have uploaded proof of their identity to the platform, hence may be perceived less anonymity according to BGCM. Perceived anonymity shows positive effects on positive attitudes towards cyberbullying, which in turn has positive effects on future cyberbullying perpetration (Barlett and Gentile, 2012). Hence, unverified users, who may be perceived more anonymity than verified users, they may be more likely to share pro-cyberbullying posts. We will test it using our dataset.

In addition, been verified is helpful for users to be more central to the information retweeting network (González-Bailón and De Domenico, 2021). For instance, verified users on Twitter, under its previous management, have been identified by Twitter as accounts of public interest (Simon et al., 2014). Public figures are likely to be verified users (Wang and Zhu, 2021). Also, verified users tend to be more active, as was illustrated in the major spike in Twitter use from verified users when China experienced its first COVID-19-related death (Chen et al., 2020). This is anticipated because public figures and news sources frequently report breaking news immediately, and verified users play a prominent role in disseminating information as messages from verified media accounts are commonly shared through retweets. (González-Bailón and De Domenico, 2021). On Weibo, only around 17% of users are verified. If a verified user shares inappropriate information, they will be suspended by the platform for a period (Zhang and Lu, 2016). For famous verified users such as celebrities, sharing inappropriate information may harm their reputation and public image (Wang et al., 2014); Therefore, this also led to the hypothesis that verified users on Weibo exercise greater caution in their online behavior regarding pro-cyberbullying content compared to unverified users.

Existing research and theories suggest a negative association between verification status and pro-cyberbullying attitudes. However, despite this association, verified users' higher activity and influence on social networks amplify the reach of their pro-cyberbullying content. The expedited dissemination of posts from verified users may be attributed to the preference for information from trusted sources, which demands less cognitive effort compared to content that contradicts established beliefs or perspectives (Knobloch-Westerwick et al., 2020). Moreover, as individuals share or repost content that could be perceived as cyberbullying, it may contribute to perpetuating such behaviors in an amplified manner among online users (Steinmetz et al., 2014). Therefore, we will examine whether or not pro-cyberbullying content from verified users on Weibo is more likely to be reposted than that from unverified users.

2.3 Emotions and cyberbullying

The increased risk of cyberbullying during the pandemic may be because people feared exposure to the virus, because of reduced income, fear of death or hospitalization, or stigmatization (Yang et al., 2022), all stressors that are related to emotional problems such as increased use of emotional expressions, depression, or anxiety (Holmes et al., 2020; Wong et al., 2020; Barlett et al., 2021b). Researchers established the framework for a General Strain Theory (GST) with a core emphasis on negative emotions and affect (Agnew and White, 1985). This theory posits that negative affective states, such as anger and related emotions, emerge in response to certain stimuli, thereby heightening the likelihood of delinquent adaptations (Mazerolle et al., 2000). GST has been extensively utilized to analyze aggressive behaviors, including cyberbullying (Lianos and McGrath, 2018). Using a sample of 1,103 Chinese adolescents survey, researchers found that adolescents facing financial strain are at a heightened risk of experiencing emotional challenges such as anger and depression, which in turn, increase the likelihood of engaging in cyberbullying perpetration (Wang and Jiang, 2023).

In addition, previous studies have found that emotions are closely related to the occurrence of cyberbullying and that the expression of emotions has gender characteristics (Li, 2006; Armenti and Babcock, 2021; Santos et al., 2021). For example, women were found to use more emotional words in comparison with men when defending their image, especially in relation to anger (Paciello et al., 2020). Anger expression is consistent with the external attribution of blame (Rico et al., 2017), which also suggests a relationship between cyberbullying and expressions of anger. Experiencing anger is linked to higher participation in cyberbullying (Wollebæk et al., 2019), possibly because individuals experiencing anger tend to respond aggressively and engage in cyberbullying more frequently (Den Hamer and Konijn, 2016). Gender-specific differences were also found in experiences of online sexuality and intimacy, and aggressive and problematic online encounters (Chang et al., 2021). In contrast to males, females, who typically possess higher levels of affective and cognitive empathy, are less inclined to engage in online aggression (Ang and Goh, 2010). Therefore, we aim to test whether in females' pro-cyberbullying posts there are less anger/affect words than in males' pro-cyberbullying posts.

Given that anonymity (verified status) and gender are recognized as influential factors in cyberbullying behavior, with affect and anger serving as potential mediators of cyberbullying attitudes, our study is also interested to examine the variations in affect and anger within pro-cyberbullying posts among verified and unverified individuals of both genders.

3 Materials and methods

This section introduces the social media data collection, preprocessing, processing (annotation and classification), emotion words extraction, statistical analysis, and robustness check.

3.1 Data collection and preprocessing

Before collecting data, we need to confirm the searching keyword that can be applied. Our goal was to analyze cyberbullying-related post related to people diagnosed with COVID-19. We used Baidu Index and Weibo search results to identify keywords and phrases. Baidu Index,[1] similar to Google Trends,[2] has been used for keyword selection and filtering (Vaughan and Chen, 2015; Fang et al., 2021). It measures the popularity and relevance of search terms in online discourse, providing a quantitative measure of their significance. Both keyword stemming and related keyword generation methods can facilitate the identification and retrieval of documents relevant to a specified keyword. Researchers used Baidu Index and identified keywords that could accurately represent terms

1 https://index.baidu.com/v2/index.html
2 https://trends.google.com/trends/

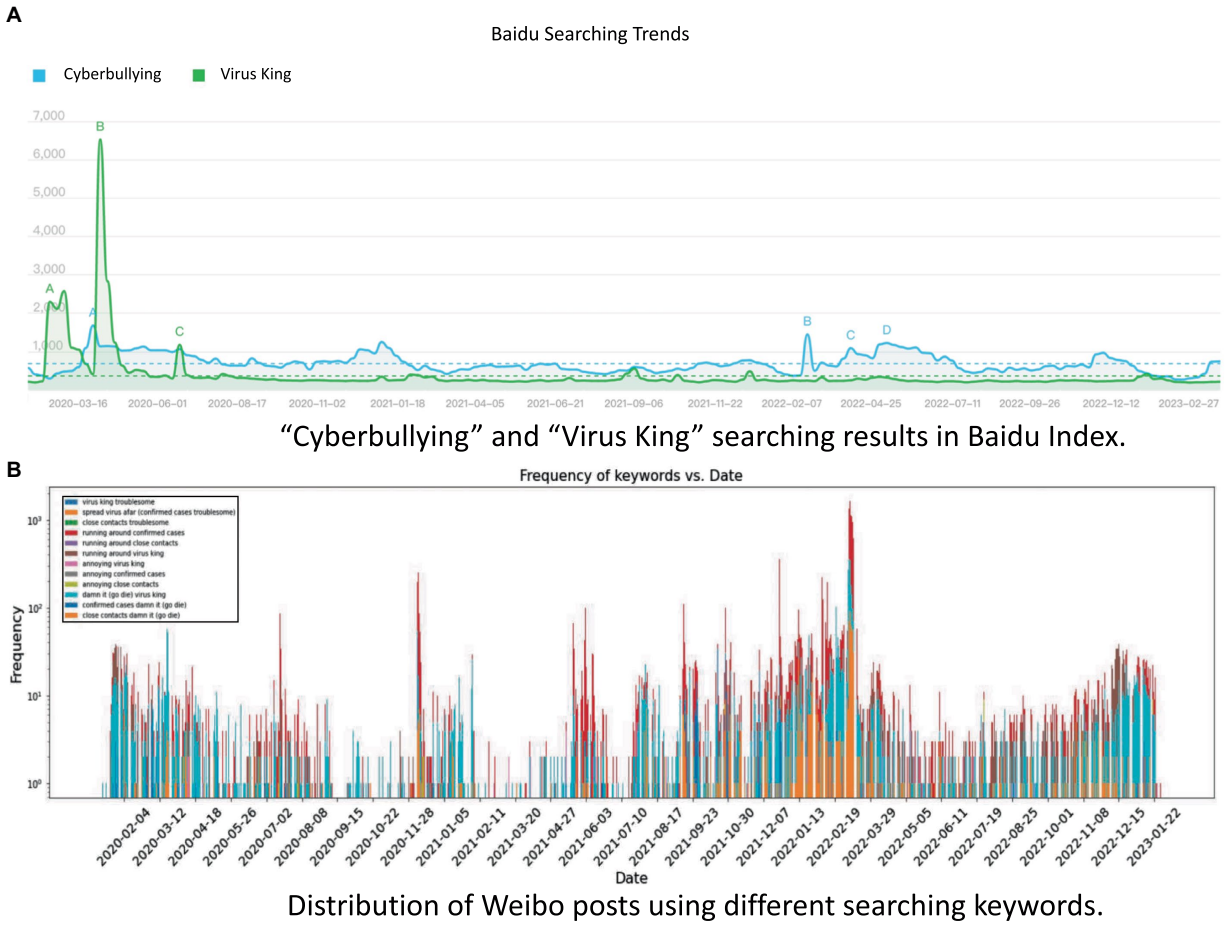

FIGURE 1
Baidu Index searching results and distribution of posts' searching results. **(A)** "Cyberbullying" and "Virus King" searching results in Baidu Index. **(B)** Distribution of Weibo posts using different searching keywords.

commonly linked with influenza epidemics (Yuan et al., 2013). We will employ the Baidu Index to identify cyberbullying-related keywords associated with COVID-19.

Firstly, we searched for "cyberbullying" and "virus king" (This search term could be translated into English as 'poison king' or 'virus king', see the searching results in Figure 1A) in Baidu Index to gain initial insights on the range of relevant terms. This is because cyberbullying is the topic we focus on and "virus king" is a term used to refer to the stigma experienced by super-spreaders of coronavirus during COVID-19 pandemic (Ting and Shamsul, 2022).

Secondly, we manually checked the top 10 monthly search results for "cyberbullying" and "virus king" on Weibo from January 1, 2020, to February 28, 2023. This step involved identifying relevant keywords with similar meanings to "cyberbullying" and "virus king" from the top 10 search results. We examined 260 posts and a panel consisting of authors 1 and 2 identified two keywords per page, specifically focusing on keywords associated with resentment (Inner Annotator Agreement of F measure = 0.91).

From this analysis, we prepared the most frequent related keywords that can be applied for collecting Weibo posts. These keywords could be translated into English as: "virus king troublesome," "spread virus afar (confirmed cases troublesome)," "close contacts troublesome," "running around confirmed cases," "running around close contacts," "running around virus king," "annoying virus king," "annoying confirmed cases," "annoying close contacts," and "damn it (go die) virus king," "confirmed cases damn it (go die)," or "close contacts damn it (go die)".[3]

Data were collected from the Weibo API[4] using keyword searching approach. We obtained in total 33,484 unique posts with at least one of the 12 keywords/phrases. For each post, the poster's username, gender, verification status, the content of the post, and the time of the post were recorded. The visualization of the availability of these 12 keywords with dates was illustrated in Figure 1B for clarification.

3.2 Data processing

We conducted two data labeling steps to define the classification tasks of interests.

3 Chinese keywords are as follows: "毒王(不省心), 千里投毒 (确诊者不省心), 密接不省心,确诊者乱跑, 密接者乱跑, 毒王乱跑, 毒王烦人,确诊者烦人, 密接烦人, 毒王要死了,确诊者要死了,密接要死了."

4 https://open.weibo.com/

(1) Identify whether the posts were cyberbullying-related or not. Cyberbullying-related or not indicate that the post is about the cyberbullying topic or not, if the 'cyberbullying-related or not' was labeled as 'yes'. Specifically, cyberbullying-related posts were defined as those which mentioned one or more of the 12 keywords and which annotators felt constituted cyberbullying directed towards COVID-19 patients. For example, "During these past few days, I got into arguments with several trolls on Weibo due to the Putian COVID-19 outbreak. Cyberbullying is extremely terrifying during a public health crisis. Nobody wants to get infected with the coronavirus, and patients who strictly follow the quarantine regulations are also innocent." This post was labeled as "True" for being cyberbullying-related. However, it was labeled "False" for the post "Shanghai has reported 4 new locally transmitted confirmed cases. We remind everyone that the epidemic is still ongoing, so please do not let your guard down. Wear masks, avoid wandering aimlessly, and refrain from unnecessary travel." In this step, we labeled 1,500 randomly selected posts for whether they were cyberbullying-related. Among them, 565 (37.7%) posts were coded as cyberbullying-related posts and the remaining 935 (62.3%) as not cyberbullying-related. The average pairwise Cohen's Kappa of the three coders was 0.72.

(2) Label the affirmed cyberbullying posts into three types: 'pro-cyberbullying', 'anti-cyberbullying', or unclear. Pro-cyberbullying posts were defined as posts/reposts with content that is harsh towards or unfairly critical of people with COVID, or those that directly called confirmed cases "virus king." For example, "these confirmed cases are so unethical. They know they have COVID-19 but still wander around recklessly." Or "Why do not those wandering confirmed COVID-19 cases just go die? They are so annoying." Anti-cyberbullying posts/reposts with content that is critical of people/content classified as cyberbullying. For example, "Regardless, cyberbullying towards confirmed cases is wrong," was labeled as "anti-cyberbullying." Unclear posts were defined as posts where we could not tell the attitudes of the poster to cyberbullying. For example, "What is your opinion about the cyberbullying of confirmed cases?" was labeled as "unclear." The average Cohen's kappa value of three coders was 0.85 (0.85, 0.81, and 0.88 separately) suggesting good agreement (Viera and Garrett, 2005; Babvey et al., 2021).

Based on previous research we applied five classification methods to the manually labeled data: k-nearest neighbors (KNN), random forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGB), Multi-Layer Perceptron (MLP, a type of neural network), and decision tree (DT) (Kotsiantis et al., 2007). A brief review of the above-mentioned machine learning models can be found in Al-Garadi et al. (2019). For each method, the bag-of-words (tokenizing data using Python package 'jieba'[5]) was applied to extract features and Term Frequency-Inverse Document Frequency (TF-IDF) was applied to assign weights to words. Performance evaluation metrics for model comparisons include positive predictive values (PPV, precision), negative predictive value (NPV), recall, and F1 score, and their confidence intervals (CI) obtained by using the bootstrapping approach (Cho et al., 1997). We then conducted two rounds of evaluations.

Firstly, we identified whether posts were cyberbullying-related or not, using 10-fold cross-validation over an internal validation dataset (randomly sampled from 90% of the labeled data). Performance comparisons were summarized in Table 1, where DT ranks the best model to predict cyberbullying posts. The hold-out performances of the DT model (using the same optimal parameters) predict the labels of the remaining 10% held-out data.

Secondly, we used these models to predict whether cyberbullying-related posts appear to be pro-cyberbullying or anti-cyberbullying (Table 2), and found that XGB achieved the best performance. And hold-out performances of XGB over the remaining 10% of labeled data. Regarding the performance of the best model (i.e., XGB) over the testing data, although binary prediction performance for the positive is not good, its prediction strength for the negative is good with (F1 = 0.79), indicating the classifier's good performance in distinguishing negative from overall posts. This satisfying the choice the best prediction model for binary classifier, as has been used and interpreted by Yang et al. (2023). As seen in the obtained performance summary for predicting 'cyberbullying-related or not' (Table 2), the best model XGB achieved high performance for predicting prevalence-dependent negative instances (F1 = 0.79, with recall = 0.93), demonstrating its ability to exclude 'cyberbullying' from overall posts with high confidence. Figure 2 shows the number of each label.

3.3 Emotion words extraction from social media posts

To extract the number of emotion-related words, we applied the Linguistic Inquiry and Word Count (LIWC) tool (Tausczik and Pennebaker, 2010), which divides words into psychologically meaningful categories and has previously been found effective in extracting emotion words from Twitter posts and online reviews (Tumasjan et al., 2010; Del Pilar Salas-Zárate et al., 2014). We applied the LIWC simplified Chinese edition (2015) and conducted text segmentation using the Python "*jieba*" package to segment Chinese words (Zhang and Goncalves, 2016). We then summarized the rates of two categories of LIWC-based emotion words in the cyberbullying-related posts: affect (all kinds of emotions) and anger.

3.4 Statistical analysis

Descriptive statistics of the extracted characteristics from texts or other sources of data, continuous variables were presented as mean and standard deviation. The standardized mean difference (SMD) was applied to compare continuous variables' differences in two targeted sets: posts pro-cyberbullying and anti-cyberbullying, within different groups of users based on their verification status and gender.

Standardized mean difference (SMD, also known as Cohen's d measure) is given by the following Equation (1) for continuous variables (Hedges et al., 2012):

$$SMD = \frac{\overline{X_1} - \overline{X_2}}{\sqrt[2]{\left(s_1^2 + s_2^2\right)/2}} \quad (1)$$

where X_1 and X_2 are sample mean for the treated and control groups, respectively; s_1^2 and s_2^2 are sample variance for the treated and control groups. It is noted that the difference between two groups is

5 https://pypi.org/project/jieba/

TABLE 1 Classifier performance for identifying whether posts are cyberbullying-related or not, using 10-fold cross-validation approach.

	Training 90% labeled data, CV = 10	10% blind data as test data			
	F1-score [95% CI]	Cyberbullying-related or not	F1-score [95% CI]	Precision [95% CI]	Recall [95% CI]
KNN classifier	0.80 (0.75, 0.85)	True	0.37 (0.15, 0.56)	0.45 (0.18, 0.73)	0.32 (0.12, 0.53)
		False	0.63 (0.48, 0.77)	0.58 (0.41, 0.74)	0.70 (0.52, 0.88)
Random forest	0.87 (0.83, 0.91)	True	0.30 (0.09, 0.51)	0.60 (0.22, 1.00)	0.21 (0.05, 0.39)
		False	0.71 (0.59, 0.83)	0.60 (0.44, 0.74)	0.89 (0.77, 1.00)
GBM	0.87 (0.83, 0.91)	True	0.41 (0.18, 0.63)	0.61 (0.30, 0.90)	0.32 (0.13, 0.56)
		False	0.71 (0.58, 0.83)	0.62 (0.47, 0.78)	0.84 (0.69, 0.96)
XGB	0.85 (0.79, 0.89)	True	0.47 (0.24, 0.67)	0.61 (0.33, 0.89)	0.39 (0.18, 0.61)
		False	0.71 (0.57, 0.83)	0.64 (0.47, 0.80)	0.81 (0.64, 0.96)
MLP	0.86 (0.82, 0.90)	True	0.24 (0.00, 0.46)	0.65 (0.00, 1.00)	0.15 (0.00, 0.32)
		False	0.72 (0.59, 0.83)	0.59 (0.44, 0.74)	0.94 (0.83, 1.00)
Decision Tree	**0.83 (0.78, 0.88)**	**True**	**0.51 (0.30, 0.71)**	**0.55 (0.31, 0.79)**	**0.49 (0.26, 0.71)**
		False	**0.66 (0.51, 0.80)**	**0.64 (0.46, 0.81)**	**0.70 (0.52, 0.87)**

Bold values mean the best performed model and its performances on training and test data.

TABLE 2 Classifier performances for whether cyberbullying-related posts appear to be pro-cyberbullying or anti-cyberbullying, using 5-fold cross-validation approach.

	Training 90% labeled data, CV = 5	10% blind data as test data, 90% data as training data			
	F1-score	Cyberbullying-related or not	F1-score [95% CI]	Precision [95% CI]	Recall [95% CI]
KNN classifier	0.75 (0.70, 0.79)	True	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
		False	0.76 (0.66, 0.86)	0.65 (0.51, 0.79)	0.94 (0.85, 1.00)
Random forest	0.95 (0.93, 0.98)	True	0.02 (0.00, 0.17)	0.19 (0.00, 1.00)	0.01 (0.00, 0.09)
		False	0.80 (0.70, 0.89)	0.66 (0.53, 0.80)	1.00 (1.00, 1.00)
GBM	0.94 (0.91, 0.97)	True	0.16 (0.00, 0.42)	0.48 (0.00, 1.00)	0.10 (0.00, 0.29)
		False	0.78 (0.67, 0.88)	0.67 (0.53, 0.80)	0.95 (0.85, 1.00)
XGB	**0.92 (0.89, 0.95)**	**True**	**0.25 (0.00, 0.50)**	**0.56 (0.00, 1.00)**	**0.17 (0.00, 0.38)**
		False	**0.79 (0.68, 0.89)**	**0.69 (0.54, 0.82)**	**0.93 (0.83, 1.00)**
MLP	0.98 (0.96, 0.99)	True	0.16 (0.00, 0.42)	0.48 (0.00, 1.00)	0.10 (0.00, 0.29)
		False	0.78 (0.67, 0.88)	0.67 (0.53, 0.80)	0.95 (0.85, 1.00)
Decision Tree	0.90 (0.87, 0.94)	True	0.24 (0.00, 0.48)	0.42 (0.00, 0.83)	0.18 (0.00, 0.39)
		False	0.76 (0.64, 0.86)	0.67 (0.53, 0.82)	0.87 (0.74, 0.97)

Bold values mean the best performed model and its performances on training and test data.

no long dependent on the unit of measurement and thus variables with different types of measurements can be compared on SMD scale. The smaller the SMD, the smaller the difference in the corresponding covariates (Zhang et al., 2019). Although a threshold value such as 0.1 is not a fixed absolute standard because there is no mathematically accurate basis, a common rule of thumb for determining no group difference considering it to be achieved when the absolute value of SMD is less than 0.1 (Sun et al., 2023; Fadini et al., 2024). That is, if a SMD value is less than 0.1, the difference between the two groups is small. In our study, we used threshold value 0.2, which is even more strictly to define the difference between group comparison.

3.5 Robustness check

To ensure the persuasive validity of the examination results for posts targeting pro-cyberbullying, we manually annotated 977 posts that had already been predicted by machine learning models (one of our authors annotated the text, followed by a thorough double-check by another to ensure uniform understanding between them). The results showed that 904 posts were still labeled as pro-cyberbullying. This implies a machine labeling accuracy of 92.5%. Further statistical analysis (SMD) was conducted on the posts manually labeled as pro-cyberbullying, and the results have been included in Table 3 (row

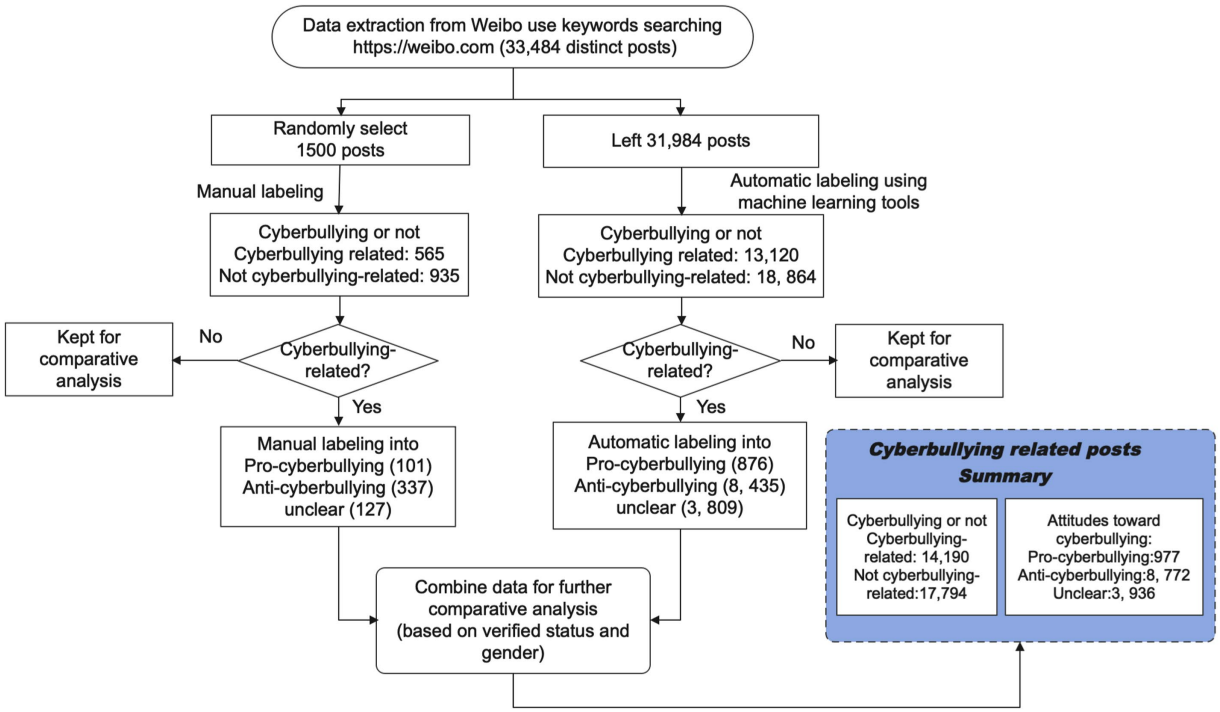

FIGURE 2
Data collection and pre-processing processes.

3 and row 5) and Table 4 (row 3 and row 5). The same findings can be observed from the obtained analysis results, showcasing the robustness of the developed machine learning models for cyberbullying prediction from social media posts, and its potential to be used in larger scale for the identification and surveillance of cyberbullying during a public health crisis. Please find the labeled data using this link.[6]

4 Results

4.1 Supportiveness of cyberbullying

The daily distributions (smoothed) of posts anti-cyberbullying and pro-cyberbullying are presented in Figure 3.

Figure 3 shows that during the early months (from January 2020 to March 2020) after the start of the COVID-19 pandemic, the number of pro-cyberbullying posts was similar to the number of anti-cyberbullying posts. Later, anti-cyberbullying posts are much more frequent than pro-cyberbullying posts.

Further, the mean value and standard deviation (SD) were calculated for the number of affect/anger words in the pro-cyberbullying posts and anti-cyberbullying posts. Table 5 summarizes the differences between the pro-cyberbullying group and the anti-cyberbullying group.

The number of pro-cyberbullying posts was around 10% of the number of anti-cyberbullying posts (see Table 5). Among the pro-cyberbullying posts, the average number of emotion words (affect) were significantly smaller than in anti-cyberbullying posts. Additionally, the average number of reposts in anti-cyberbullying posts was slightly larger than that in pro-cyberbullying posts (although the differences are not significant) according to the SMD threshold (i.e., SMD 0.2). The proportion of pro-cyberbullying posts from females is 1.36 times the number of posts from males, and the proportion of anti-cyberbullying posts from females is 0.96 times the number of posts from males.

4.2 Association of verification status and gender with emotion expressions in the pro-cyberbullying posts and anti-cyberbullying posts

We analyzed the differences of overall emotions (affect) and anger within pro-and anti-cyberbullying posts, the data analysis results are summarized in Table 3.

Table 3 shows that unverified male users used significantly more anger words than verified male users in pro-cyberbullying posts (SMD > 0.2 and the result is robust), while unverified male users have more overall emotion and anger expression than verified male users in anti-cyberbullying posts. Verified female users express significantly more anger (but not overall emotion) than verified male users in pro-cyberbullying posts (SMD > 0.2), while they express significantly more overall emotion than verified male users in anti-cyberbullying posts (SMD > 0.2). No significant difference shows in overall emotion and anger expression between verified female users and unverified female users (SMD < 0.2). These imply that unverified male users and verified female users were more likely to share pro-cyberbullying posts during the COVID-19 pandemic.

6 https://github.com/sunshinegirl5566/pro-cyberbullying-data/tree/main

TABLE 3 Summary of mean differences of emotion words frequencies from users of different gender and verified types in pro-cyberbullying and anti-cyberbullying posts.

	Emotional words frequency	Male verified Mean (SD); *N*; (1)	Male unverified Mean (SD); *N*; (2)	Female verified Mean (SD); *N*; (3)	Female unverified Mean (SD); *N*; (4)	SMD (1)–(2)	SMD (3)–(4)	SMD (1)–(3)	SMD (2)–(4)
Pro-cyberbullying	Affect	5.6 (9.3); $n = 61$	6.4 (7.6); $n = 186$	4.8 (7.6);$n = 129$	5.5 (7.4); $n = 601$	0.1	0.01	0.12	0.12
	Affect (labeled)	6.0 (9.5); $n = 57$	6.8 (7.6); $n = 176$	5.2 (7.8);$n = 117$	5.8 (7.6); $n = 554$	0.09	0.08	0.09	0.12
	Anger	0.8 (2.5); $n = 61$	1.8 (4.2); $n = 186$	1.4 (3.4); $n = 129$	1.4 (4.1); $n = 601$	0.29*	0.09	0.21*	0.09
	Anger (labeled)	0.8 (2.6); $n = 57$	1.8 (4.3); $n = 176$	1.5 (3.5); $n = 117$	1.5 (4.2); $n = 554$	0.29*	<0.01	0.22*	0.09
Anti-cyberbullying	Affect	6.6 (6.7); $n = 891$	8.4 (7.5); $n = 1791$	8.0 (7.6); $n = 1,107$	9.2 (8.0); $n = 4,983$	0.25*	0.15	0.19	0.11
	Anger	2.0 (3.0); $n = 891$	1.8 (3.1); $n = 1791$	2.0 (3.3); $n = 1,107$	2.0 (3.2); $n = 4,983$	0.07	0.01	<0.01	0.06

*Indicates SMD > 0.2.

TABLE 4 Summary of the posts from users of different gender and verification status.

		Male verified Mean (SD); (1)	Male unverified Mean (SD) (2)	Female verified Mean (SD) (3)	Female unverified Mean (SD) (4)	SMD (*p*-value) (1)–(2)	SMD (P-value) (3)–(4)	SMD (P-value) (1)–(3)	SMD (P-value) (2)–(4)
Pro-cyberbullying	Number of reposts	4.6 (12.5)	0.6 (2.8)	0.5 (2.6)	0.4 (4.9)	0.45*	0.03	0.48*	0.06
	Number of reposts (labeled)	4.8 (12.9)	0.6 (2.9)	0.5 (2.7)	0.2 (1.4)	0.45*	0.14	0.47*	0.20*
	Number of posts	61	186	129	601				
	Number of posts (labeled)	57	176	117	554				
Anti-cyberbullying	Number of reposts	134.0 (2280.1)	1.51 (23.14)	35.6 (356.2)	1.52 (22.71)	0.08	0.13	0.06	<0.01
	Number of posts	891	1791	1,107	4,983	/	/	/	/

*Indicates SMD > 0.2.

10.3389/fpsyg.2024.1395668

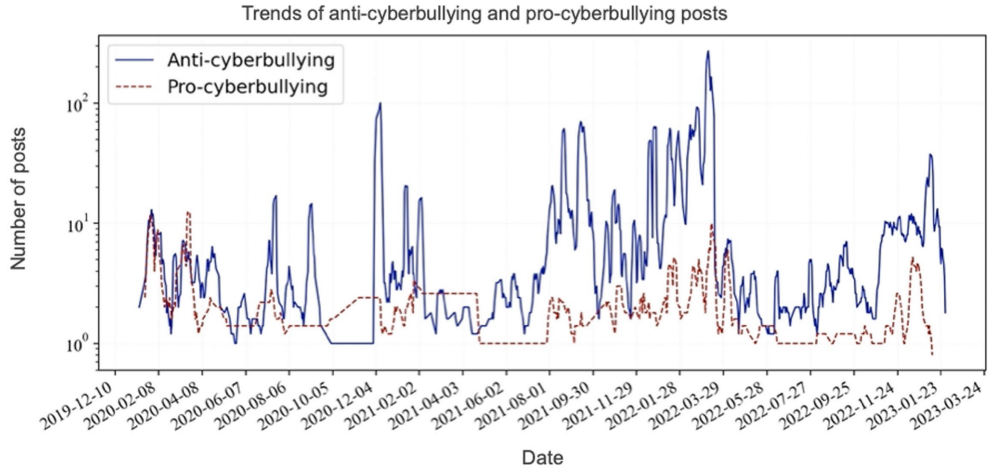

FIGURE 3
Temporal trends in the number of anti-cyberbullying and pro-cyberbullying posts, 2020 to 2023.

TABLE 5 Comparison of pro-cyberbullying and anti-cyberbullying posts.

Characteristics	Pro-cyberbullying Mean (SD); (*N* = 977)	Anti-cyberbullying Mean (SD); (*N* = 8,772)	Standardized Mean Difference (SMD)
Affect (all kinds of emotional words, e.g., happy, sad, disgust etc.)	5.6 (7.6)	8.4 (7.7)	0.39*
Anger (mad, hate, kill etc.)	1.4 (3.9)	2.0 (3.1)	0.14
Number of reposts	0.7 (5.3)	19.3 (738.6)	0.04
Gender	male:247; female: 730	male:2682; female: 6090	/
Gender ratio	1.36	0.96	

*Indicates SMD > 0.2. Gender ratio for pro-cyberbullying posts = (Female user number support cyberbullying/Total number of female users)/(Male user number support cyberbullying/Total number of male users).

4.3 Differences in the reposts number of anti-cyberbullying and pro-cyberbullying posts

Table 4 summarizes pair-wise comparisons among several groups: verified male users as Type (1), unverified male users as Type (2), verified female users as Type (3), and unverified females as Type (4).

Table 4 shows that the number of reposts from verified male users was significantly larger than that from unverified male users and verified female users in pro-cyberbullying posts (SMD > 0.2). However, no significant differences were evident in the number of reposts between each pair of user groups in anti-cyberbullying posts (SMD < 0.2).

5 Discussion

5.1 Findings and practical implications

This study examined the pro-cyberbullying and anti-cyberbullying information sharing behavior of social media users regarding COVID-19 patients. We found that while there were a lot of posts about cyberbullying, the posts that could be considered pro-cyberbullying were in the minority (one in ten). Though only a small proportion of posts were pro-cyberbullying, this study suggests that the harmfulness of cyberbullying could be amplified, reflected in the large number of anger words in pro-cyberbullying posts from female verified users than the pro-cyberbullying posts from male verified users, as well as the larger number of female users in Weibo (Sheng et al., 2022). Practically, more attention should be paid to cyberbullying posts from verified male users because their posts had a significantly larger number of reposts than those from other users including male unverified users, and female users (verified or unverified).

Further, among all the pro-cyberbullying posts (977), 787 (80.5%) posts were from unverified users and 190 (19.5%) posts were from verified users. There are 4.14 times of pro-cyberbullying posts of the unverified users than that of the verified users. This study also found that users varied in expressing general emotions (affect) or anger in both pro-cyberbullying and anti-cyberbullying posts in accordance with their gender and verification status. Compared to unverified male users, verified male users have higher public visibility (verified users are more likely to be public figures) and may be more careful about maintaining a positive image. Expressing anger towards vulnerable groups is more likely to damage the image of verified users (Wang and Zhu, 2021).

Finally, expressing anger may be reasonable and could benefit people's image in some circumstances. For instance, scapegoating strategy suggests that people/organizations might seek to shift blame to others (Coombs, 2015). Specifically, if individuals observe a negative outcome of others' behaviors and attribute responsibility to

him/her, they may feel anger and have negative feelings (Schwarz, 2012). In the anti-cyberbullying posts, female users (especially verified female users) expressed more emotions than male users (both verified and unverified), following social norms for emotional expression (Simon and Nath, 2004).

5.2 Theoretical implications

We found gender differences in sharing pro-cyberbullying posts of COVID-19 patients. The female to male ratio of 1.19 to 1 for pro-cyberbullying posts is slightly larger than the ratio of 0.93 for posts that are anti-cyberbullying. These findings may indicate that males are more cautious in sharing pro-cyberbullying posts than females. This discovery diverges from previous research, which, guided by self-control theory, suggested that females might be less inclined to engage in cyberbullying. However, it becomes reasonable when considering the perspective of the cyberbullying perpetrator, who deems others' behavior as immoral. During the manual validation of pro-cyberbullying posts, it was evident that individuals who are pro-cyberbullying felt angered by what they perceived as the immoral actions of confirmed COVID-19 cases who ventured outdoors without wearing masks, thus endangering others. This phenomenon may be elucidated by Bandura's moral agency theory (Kurtines and Gewirtz, 1991), which introduces the concept of moral disengagement—a cognitive process allowing individuals to act in a manner contrary to their personal and societal norms (Kurtines and Gewirtz, 1991). Therefore, we argue that in characterizing cyberbullying behavior through the lens of self-control theory, integrating moral agency theory is imperative.

Larger portion of pro-cyberbullying posts originated from unverified users rather than verified ones, lending support to the BCGM hypothesis, which proposes that perceived anonymity correlates positively with positive attitudes toward cyberbullying (Barlett et al., 2017). This discovery underscores the validity of using verification status to gauge the perceived anonymity of social media users. In addition, in the pro-cyberbullying posts, female users expressed more anger than male users (both verified and unverified). This finding suggests the potential of further considering gender as a moderating factor in the GST theory. In addition, compared with the verified male users, the unverified male users expressed more anger in pro-cyberbullying posts. This is reasonable, because the emotional responses of individuals are highly related to individuals' cognitive processes (Armenti and Babcock, 2021), and people are cognitively inspired once they express anger at injustice and unethical phenomena. Verified females contradicting social norms and expressing more anger than verified males could also be understood through the lens of moral agency theory. This theory posits that individuals may justify actions that defy both their personal and societal norms, thus explaining the observed behavior (Kurtines and Gewirtz, 1991).

In summary, this study offers a new way to validate current theories on cyberbullying using social media data. By extracting the gender of social media users, we were able to test the self-control theory by analyzing the sharing behavior of males and females regarding pro-cyberbullying posts. Additionally, by examining the anonymity of social media users through their verification status, this study provides a new perspective for validating the BGCM. Through analyzing the ratio of anger/affect words in pro-cyberbullying posts among male/female users and verified/unverified users, this research has the potential to contribute to the development of new theories related to cyberbullying on social media.

5.3 Limitations and future research

However, this study had a number of limitations. Firstly, our research is based on Weibo data. Due to limitations with its open Application Programming Interface, we were unable to obtain a complete set of relevant data. Instead, we could only access a portion of randomly selected data and the sample size of pro-cyberbullying posts was relatively small compare with anti-cyberbullying posts. This implies that our research findings need further validation in larger datasets. Secondly, the keywords we choose are only representative keywords, but not a complete set of keywords which needs further studies to fill this gap. Thirdly, when categorizing the data related to cyberbullying, pro-cyberbullying, and anti-cyberbullying, we employed supervised machine learning techniques. However, the size of our testing set was small (e.g., in the second-round classification, the test set size is around 100, which may lead to the number of pro-cyberbullying posts is around 20), which suggests that there is need to enlarge of the dataset we now applied in the future. Fourthly, although efforts were made to ensure accurate translation, potential limitations inherent in the translation process from Chinese to English, such as cultural references and idiomatic expressions that may not fully convey the original intent or meaning. Lastly, the annotation set we used for identifying posts pro-or anti-cyberbullying was not extensive. Therefore, it is necessary for further research to expand such datasets and conduct research on them.

6 Conclusion

The rapid expansion of social media platforms over the past decade increased the prevalence of cyberbullying victimization and perpetration. Conducting investigations into cyberbullying is of great significance. The findings of this study provide gender-, verification status-, and emotion-specific empirical evidence about attitudes to cyberbullying in social media during a major public health crisis, which could be used to improve cyberbullying-post-identification algorithms. The study insights could also be employed by government agencies to mitigate the negative effects of online cyberbullying behaviors.

Data availability statement

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

Author contributions

LL: Conceptualization, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. JZ: Methodology, Validation, Writing – review & editing. SM: Supervision, Validation, Writing – review &

editing, Funding acquisition. RS: Supervision, Validation, Writing – review & editing, Funding acquisition. AR: Funding acquisition, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. RS is part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London. AR and RS are supported by Health Data Research UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. RS is additionally part-funded by the NIHR Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust, and by the DATAMIND HDR UK Mental Health Data Hub (MRC grant MR/W014386). LL is part funded by the National Natural Science Foundation of China under Grant (72104080). This research was supported by the UK Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), which is funded by the British Heart Foundation, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health and Care Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation, and Wellcome.

Conflict of interest

RS declares research support received in the last 3 years from Janssen, Takeda, and GSK. The research reported in this article did not receive any funding from the companies listed.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The views expressed in this Article are those of the authors and not necessarily those of the UK Prevention Research Partnership or any other funder.

References

Agnew, R., and White, H. R. (1985). *An empirical test of general strain theory*. New York: Greenberg.

Ak, Ş., Özdemir, Y., and Kuzucu, Y. (2015). Cybervictimization and cyberbullying: the mediating role of anger, don't anger me! *Comput. Human Behav.* 49, 437–443. doi: 10.1016/j.chb.2015.03.030

Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., et al. (2019). Predicting cyberbullying on social Media in the big Data era Using Machine Learning Algorithms: review of literature and open challenges. *IEEE Access* 7, 70701–70718. doi: 10.1109/ACCESS.2019.2918354

Ang, R. P., and Goh, D. H. (2010). Cyberbullying among adolescents: the role of affective and cognitive empathy, and gender. *Child Psychiatry Hum. Dev.* 41, 387–397. doi: 10.1007/s10578-010-0176-3

Armenti, N. A., and Babcock, J. C. (2021). Borderline personality features, anger, and intimate partner violence: an experimental manipulation of rejection. *J. Interpers. Violence* 36, NP3104–NP3129. doi: 10.1177/0886260518771686

Babvey, P., Capela, F., Cappa, C., Lipizzi, C., Petrowski, N., and Ramirez-Marquez, J. (2021). Using social media data for assessing children's exposure to violence during the COVID-19 pandemic. *Child Abuse Negl.* 116:104747. doi: 10.1016/j.chiabu.2020.104747

Bansal, S., Garg, N., Singh, J., and Van Der Walt, F. (2023). Cyberbullying and mental health: past, present and future. *Front. Psychol.* 14:1279234. doi: 10.3389/fpsyg.2023.1279234

Barlett, C. P. (2017). From theory to practice: cyberbullying theory and its application to intervention. *Comput. Human Behav.* 72, 269–275. doi: 10.1016/j.chb.2017.02.060

Barlett, C. P., Bennardi, C., Williams, S., and Zlupko, T. (2021a). Theoretically predicting cyberbullying perpetration in youth with the BGCM: unique challenges and promising research opportunities. *Front. Psychol.* 12, 1–8. doi: 10.3389/fpsyg.2021.708277

Barlett, C., Chamberlin, K., and Witkower, Z. (2017). Predicting cyberbullying perpetration in emerging adults: a theoretical test of the Barlett Gentile cyberbullying model. *Aggress. Behav.* 43, 147–154. doi: 10.1002/ab.21670

Barlett, C. P., and Gentile, D. A. (2012). Attacking others online: the formation of cyberbullying in late adolescence. *Psychol. Pop. Media Cult.* 1, 123–135. doi: 10.1037/a0028113

Barlett, C. P., Rinker, A., and Roth, B. (2021b). Cyberbullying perpetration in the COVID-19 era: an application of general strain theory. *J. Soc. Psychol.* 161, 466–476. doi: 10.1080/00224545.2021.1883503

Barlett, C. P., Simmers, M. M., Roth, B., and Gentile, D. (2021c). Comparing cyberbullying prevalence and process before and during the COVID-19 pandemic. *J. Soc. Psychol.* 161, 408–418. doi: 10.1080/00224545.2021.1918619

Beran, T., and Qing, L. I. (2005). Cyber-harassment: a study of a new method for an old behavior. *J. Educ. Comput. Res.* 32, 265–277. doi: 10.2190/8YQM-B04H-PG4D-BLLH

Chang, Q., Xing, J., Chang, R., Ip, P., Yee-Tak Fong, D., Fan, S., et al. (2021). Online sexual exposure, cyberbullying victimization and suicidal ideation among Hong Kong adolescents: moderating effects of gender and sexual orientation. *Psychiatry Res. Commun.* 1:100003. doi: 10.1016/j.psycom.2021.100003

Chen, E., Lerman, K., and Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus twitter data set. *JMIR Public Health Surveill.* 6:e19273. doi: 10.2196/19273

Cheng, C., Lau, Y. C., and Luk, J. W. (2020). Social capital–accrual, escape-from-self, and time-displacement effects of internet use during the COVID-19 stay-at-home period: prospective, quantitative survey study. *J. Med. Internet Res.* 22, e22740–e22712. doi: 10.2196/22740

Cho, K., Meer, P., and Cabrera, J. (1997). Performance assessment through bootstrap. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 1185–1198. doi: 10.1109/34.632979

Coombs, W. T. (2015). The value of communication during a crisis: insights from strategic communication research. *Bus. Horiz.* 58, 141–148. doi: 10.1016/j.bushor.2014.10.003

Del Pilar Salas-Zárate, M., López-López, E., Valencia-García, R., Aussenac-Gilles, N., Almela, Á., and Alor-Hernández, G. (2014). A study on LIWC categories for opinion mining in Spanish reviews. *J. Inf. Sci.* 40, 749–760. doi: 10.1177/0165551514547842

Den Hamer, A. H., and Konijn, E. A. (2016). Can emotion regulation serve as a tool in combating cyberbullying? *Pers Individ Dif* 102, 1–6. doi: 10.1016/j.paid.2016.06.033

Fadini, G. P., Longato, E., Morieri, M. L., Del Prato, S., Avogaro, A., Solini, A., et al. (2024). Long-term benefits of dapagliflozin on renal outcomes of type 2 diabetes under routine care: a comparative effectiveness study on propensity score matched cohorts at

low renal risk. *Lancet Regional Health-Europe* 38:100847. doi: 10.1016/j.lanepe.2024.100847

Fang, J., Zhang, X., Tong, Y., Xia, Y., Liu, H., and Wu, K. (2021). Baidu index and COVID-19 epidemic forecast: evidence from China. *Front. Public Health* 9:685141. doi: 10.3389/fpubh.2021.685141

Kurtines, W. M., and Gewirtz, J. L. (1991). Handbook of moral behavior and development. *Vol. 1. Theory; Vol. 2. Research; Vol. 3. Application*. Lawrence Erlbaum Associates, Inc.

González-Bailón, S., and De Domenico, M. (2021). Bots are less central than verified accounts during contentious political events. *Proc. Natl. Acad. Sci. USA* 118:e2013443118. doi: 10.1073/pnas.2013443118

Görzig, A., and Ólafsson, K. (2013). What makes a bully a cyberbully? Unravelling the characteristics of cyberbullies across twenty-five European countries. *J. Child. Media* 7, 9–27. doi: 10.1080/17482798.2012.739756

Griezel, L., Finger, L. R., Bodkin-Andrews, G. H., Craven, R. G., and Yeung, A. S. (2012). Uncovering the structure of and gender and developmental differences in cyber bullying. *J. Educ. Res.* 105, 442–455. doi: 10.1080/00220671.2011.629692

Hedges, L. V., Pustejovsky, J. E., and Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Res. Synth. Methods* 3, 224–239. doi: 10.1002/jrsm.1052

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

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., et al. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* 7, 547–560. doi: 10.1016/S2215-0366(20)30168-1

Jatmiko, M. I., Syukron, M., and Mekarsari, Y. (2020). Covid-19, harassment and social media: a study of gender-based violence facilitated by technology during the pandemic. *J. Society Media* 4:319. doi: 10.26740/jsm.v4n2.p319-347

John, A., Glendenning, A. C., Marchant, A., Montgomery, P., Stewart, A., Wood, S., et al. (2018). Self-harm, suicidal behaviours, and cyberbullying in children and young people: systematic review. *J. Med. Internet Res.* 20:e129. doi: 10.2196/jmir.9044

Knobloch-Westerwick, S., Mothes, C., and Polavin, N. (2020). Confirmation Bias, Ingroup Bias, and negativity Bias in selective exposure to political information. *Commun. Res.* 47, 104–124. doi: 10.1177/0093650217719596

Kochanska, G., Murray, K., Tanya, Y., Koenigr, A. L., and Kimberly, A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Dev.* 67, 490–507. doi: 10.2307/1131828

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: a review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* 160, 3–24. doi: 10.5555/1566770.1566773

Li, Q. (2006). Cyberbullying in schools: a research of gender differences. *Sch. Psychol. Int.* 27, 157–170. doi: 10.1177/0143034306064547

Li, W., and Peng, H. (2022). The impact of strain, constraints, and morality on different cyberbullying roles: a partial test of Agnew's general strain theory. *Front. Psychol.* 13, 1–20. doi: 10.3389/fpsyg.2022.980669

Lian, Y., Zhou, Y., Lian, X., and Dong, X. (2022). Cyber violence caused by the disclosure of route information during the COVID-19 pandemic. *Humanit. Soc. Sci. Commun.* 9:417. doi: 10.1057/s41599-022-01450-8

Lianos, H., and McGrath, A. (2018). Can the general theory of crime and general strain theory explain cyberbullying perpetration? *Crime Delinq.* 64, 674–700. doi: 10.1177/0011128717714204

Machimbarrena, J. M., Calvete, E., Fernández-González, L., Álvarez-Bardón, A., Álvarez-Fernández, L., and González-Cabrera, J. (2018). Internet risks: an overview of victimization in cyberbullying, cyber dating abuse, sexting, online grooming and problematic internet use. *Int. J. Environ. Res. Public Health* 15:471. doi: 10.3390/ijerph15112471

Marr, K. L., and Duell, M. N. (2021). Cyberbullying and cybervictimization: does gender matter? *Psychol. Rep.* 124, 577–595. doi: 10.1177/0033294120916868

Mazerolle, P., Burton Jr, V. S., Cullen, F. T., Evans, T. D., and Payne, G. L. (2000). Strain, anger, and delinquent adaptations specifying general strain theory. *J. Crim. Just.* 28, 89–101.

Morales-Arjona, I., Pastor-Moreno, G., Ruiz-Pérez, I., Sordo, L., and Henares-Montiel, J. (2022). Characterization of cyberbullying victimization and perpetration before and during the COVID-19 pandemic in Spain. *Cyberpsychol. Behav. Soc. Netw.* 25, 733–743. doi: 10.1089/cyber.2022.0041

Paciello, M., Tramontano, C., Nocentini, A., Fida, R., and Menesini, E. (2020). The role of traditional and online moral disengagement on cyberbullying: do externalising problems make any difference? *Comput. Human Behav.* 103, 190–198. doi: 10.1016/j.chb.2019.09.024

Patel, M. S. (2021). Text-message nudges encourage COVID vaccination. *Nature* 597, 336–337. doi: 10.1038/d41586-021-02043-2

Rico, G., Guinjoan, M., and Anduiza, E. (2017). The emotional underpinnings of populism: how anger and fear affect populist attitudes. *Swiss Polit. Sci. Rev.* 23, 444–461. doi: 10.1111/spsr.12261

Santos, D., Mateos-Pérez, E., Cantero, M., and Gámez-Guadix, M. (2021). Cyberbullying in adolescents: resilience as a protective factor of mental health outcomes. *Cyberpsychol. Behav. Soc. Netw.* 24, 414–420. doi: 10.1089/cyber.2020.0337

Schwarz, A. (2012). How publics use social media to respond to blame games in crisis communication: the love parade tragedy in Duisburg 2010. *Public Relat. Rev.* 38, 430–437. doi: 10.1016/j.pubrev.2012.01.009

Sheng, Q., Cao, J., Bernard, H. R., Shu, K., Li, J., and Liu, H. (2022). Characterizing multi-domain false news and underlying user effects on Chinese Weibo. *Inf. Process. Manag.* 59:102959. doi: 10.1016/j.ipm.2022.102959

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., and Adini, B. (2014). Twitter in the cross fire – the use of social media in the Westgate mall terror attack in Kenya. *PLoS One* 9:e104136. doi: 10.1371/journal.pone.0104136

Simon, R. W., and Nath, L. E. (2004). Gender and emotion in the United States: do men and women differ in self-reports of feelings and expressive behavior? *Am. J. Sociol.* 109, 1137–1176. doi: 10.1086/382111

Steinmetz, J., Bosak, J., Sczesny, S., and Eagly, A. H. (2014). Social role effects on gender stereotyping in Germany and Japan. *Asian J. Soc. Psychol.* 17, 52–60. doi: 10.1111/ajsp.12044

Stults, B. J., and You, M. (2022). Self-control, cyberbullying, and the moderating effect of opportunity. *Deviant Behav.* 43, 1267–1284. doi: 10.1080/01639625.2021.1985928

Sun, Y., Jin, L., Dian, Y., Shen, M., Zeng, F., Chen, X., et al. (2023). Oral Azvudine for hospitalised patients with COVID-19 and pre-existing conditions: a retrospective cohort study. *EClinicalMedicine* 59:101981. doi: 10.1016/j.eclinm.2023.101981

Tausczik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* 29, 24–54. doi: 10.1177/0261927X09351676

Ting, S.-H., and Shamsul, M. H. (2022). Stigma-marking of COVID-19 patients in Facebook and twitter of youth in Malaysia in 2020–2021. *Youth* 2, 717–732. doi: 10.3390/youth2040051

Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. (2010). Predicting elections with twitter: what 140 characters reveal about political sentiment. *Proc. Fourth Int. AAAI Conf. Weblogs Social Media* 280, 33411–33418. doi: 10.1074/jbc.M501708200

Vaughan, L., and Chen, Y. (2015). Data mining from web search queries: a comparison of google trends and baidu index. *J. Assoc. Inf. Sci. Technol.* 66, 13–22. doi: 10.1002/asi.23201

Viera, A. J., and Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37, 360–363

Wang, L., and Jiang, S. (2023). The effects of strain and negative emotions on adolescent cyberbullying perpetration: an empirical test of general strain theory. *Curr. Psychol.* 42, 11439–11449. doi: 10.1007/s12144-021-02426-8

Wang, N., She, J., and Chen, J. (2014). How "big vs" dominate Chinese microblog: A comparison of verified and unverified users on Sina Weibo. Web Sci 2014. Proceedings of the 2014 ACM web science conference. 182–186.

Wang, M. J., Yogeeswaran, K., Andrews, N. P., Hawi, D. R., and Sibley, C. G. (2019). How common is cyberbullying among adults? Exploring gender, ethnic, and age differences in the prevalence of cyberbullying. *Cyberpsychol. Behav. Soc. Netw.* 22, 736–741. doi: 10.1089/cyber.2019.0146

Wang, C. J., and Zhu, J. J. H. (2021). Jumping over the network threshold of information diffusion: testing the threshold hypothesis of social influence. *Internet Res.* 31, 1677–1694. doi: 10.1108/INTR-08-2019-0313

Wollebæk, D., Karlsen, R., Steen-Johnsen, K., and Enjolras, B. (2019). Anger, fear, and Echo chambers: the emotional basis for online behavior. *Soc. Media Soc.* 5:205630511982985. doi: 10.1177/2056305119829859

Wong, R. Y. M., Cheung, C. M. K., and Xiao, B. (2018). Does gender matter in cyberbullying perpetration? An empirical investigation. *Comput. Human Behav.* 79, 247–257. doi: 10.1016/j.chb.2017.10.022

Wong, A. H., Roppolo, L. P., Chang, B. P., Yonkers, K. A., Wilson, M. P., Powsner, S., et al. (2020). Management of agitation during the COVID-19 pandemic. *Western J. Emerg. Med.* 21, 795–800. doi: 10.5811/westjem.2020.5.47789

Wu, W., Yu, L., Cao, X., Guo, Z., Long, Q., Zhao, X., et al. (2023). The latent profile of self-control among high school students and its relationship with gender and cyberbullying. *Curr. Psychol.* 42, 29650–29660. doi: 10.1007/s12144-022-03995-y

Yang, J., Soltan, A. A. S., Eyre, D. W., and Clifton, D. A. (2023). Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. *Nat. Mach. Intell.* 5, 884–894. doi: 10.1038/s42256-023-00697-3

Yang, F., Sun, J., Li, J., and Lyu, S. (2022). Coping strategies, stigmatizing attitude, and cyberbullying among Chinese college students during the COVID-19 lockdown. *Curr. Psychol.* 43, 8394–8402. doi: 10.1007/s12144-022-02874-w

Yuan, Q., Nsoesie, E. O., Lv, B., Peng, G., Chunara, R., and Brownstein, J. S. (2013). Monitoring influenza epidemics in China with search query from Baidu. *PLoS One* 8:e64323. doi: 10.1371/journal.pone.0064323

Yudes, C., Rey, L., and Extremera, N. (2021). The moderating effect of emotional intelligence on problematic internet use and cyberbullying perpetration among adolescents: gender differences. *Psychol. Rep.* 125, 2902–2921. doi: 10.1177/00332941211031792

Zhang, Q., and Goncalves, B. (2016). Topical differences between Chinese language twitter and Sina Weibo. Proceedings of the 25th International Conference Companion on World Wide Web. 625–628.

Zhang, Z., Kim, H. J., Lonjon, G., and Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Ann. Transl. Med.* 7:16. doi: 10.21037/atm.2018.12.10

Zhang, Y., and Lu, J. (2016). Discover millions of fake followers in Weibo. *Soc. Netw. Anal. Min.* 6, 1–15. doi: 10.1007/s13278-016-0324-2