

Original Paper

# Comparison of Mask-Wearing Behavior on Social Media and Its Relationship With Demographic Characteristics During the COVID-19 Pandemic: Social Media Data Analysis Between the United States and Japan

Kiki Ferawati<sup>1,2</sup>, PhD; Wan Jou She<sup>1,3</sup>, PhD; Shoko Wakamiya<sup>1</sup>, PhD; Eiji Aramaki<sup>1</sup>, PhD

<sup>1</sup>Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan

<sup>2</sup>Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Surakarta, Indonesia

<sup>3</sup>Faculty of Information and Human Sciences, Kyoto Institute of Technology, Kyoto, Japan

## Corresponding Author:

Eiji Aramaki, PhD

Graduate School of Science and Technology

Nara Institute of Science and Technology

8916-5 Takayama-cho

Ikoma 630-0192

Japan

Phone: 81 743-72-5250

Email: [aramaki@is.naist.jp](mailto:aramaki@is.naist.jp)

## Abstract

**Background:** Social media is one of the most accessible and extensive sources of data for tracking and understanding public reactions to COVID-19 policies. Cultural differences between the United States and Japan have resulted in highly distinctive policies and public reactions in each country.

**Objective:** This study aims to analyze the public opinions surrounding COVID-19 mask mandate through 1,102,876 and 560,873 geo-tagged tweets from the United States and Japan during the period from 2020 to 2022. We conducted 3 stages of analysis—relevance to COVID-19 masks, stance for or against masking, and whether the tweets indicate users wearing masks—to understand individuals' stance towards the mask mandate and their actual mask-wearing behavior.

**Methods:** We adopted a semisupervised approach to enhance BERT (Bidirectional Encoder Representations from Transformers) classification results due to data imbalance, which were then visualized through time series and map representations.

**Results:** In the United States, our data showed that individuals with a bachelor's degree or higher, as well as those living in states with higher household incomes, are positively correlated with positive attitudes toward mask-wearing. In contrast, in Japan, those with higher education levels or individuals aged 65 years and older were positively correlated with tweets categorized as having a stance against the mask mandate. Key events in Japan, such as the announcement of the state of emergency and the Olympics, served as major triggers for the number boost in public opinion.

**Conclusions:** Our analysis of over 1.6 million tweets from the United States and Japan revealed that public opinion shifted notably in response to major events and policy changes during the COVID-19 pandemic. While some trends align with previous research, correlations with education, age, and income suggest that social media data may reflect underlying societal divisions and algorithm-driven biases.

*JMIR Hum Factors* 2026;13:e78236; doi: [10.2196/78236](https://doi.org/10.2196/78236)

**Keywords:** COVID-19; mask-wearing; cultural differences; classification; map; correlation

## Introduction

Countries around the world have adopted a diverse strategy for managing the COVID-19 pandemic, as summarized in

the government responses index in 2020 when COVID-19 first impacted the globe [1]. The implementation of different types of policies by countries is influenced by various factors, including demographics and geographical considerations.

Among the measures taken to curb the spread of the disease, the mask mandate is recognized as one of the most widely adopted policies in several countries. Early in the pandemic, the Centers for Disease Control and Prevention (CDC) advised people in the United States to wear masks as a preventive measure [2], which was also adopted in a few other countries [3]. While the Japanese government did not mandate mask-wearing, its population had already been accustomed to wearing masks even before the COVID-19 outbreak, during the pandemic, and after [4].

The comparison between the United States and Japan is a frequently studied topic among researchers due to cultural contrasts and variations in Western and Eastern societal conventions [5,6]. Cultural views also contributed to how people in the 2 countries reacted differently as a response to health policies during the pandemic [7]. During the COVID-19 outbreak, both countries displayed contrasting attitudes regarding policy creation to control the spread of the disease and adhere to mask mandates. Undoubtedly, the vast number of social media posts is often viewed as a valuable resource for comprehending, analyzing, and even informing policymakers about societal perceptions or attitudes toward an occurrence. However, a significant obstacle for cross-cultural comparison studies is handling data from multilingual social media data. From the process of data labeling, classification, and processing, the differences between the characteristics of the 2 languages might provide interesting results to be analyzed.

Social media serves as a cost-effective and accessible alternative for assessing public responses in addition to surveys [8]. Changes in people's behavior during pandemics can be observed on social media platforms like X (formerly Twitter), as demonstrated during the 2009 H1N1 pandemic [8,9] and the COVID-19 pandemic [10]. X reflects real-time insights into individuals' reactions [11,12], making it a convenient yet highly sensitive tool for assessing the risk associated with events or interventions that may have downstream effects on policy-making and implementation [13,14]. Social media might also show declining sentiment on topics related to COVID-19 restrictions throughout the pandemic, as described in the results of a study using tweets in the 3 cities in the United States [15]. Other than previously mentioned studies, the content from social media posts, such as tweets, can also be used to assess possible outbreaks by using tools such as large language models [16].

The debate surrounding masking has generated one of the most widely discussed topics throughout the pandemic, which could be clearly traced and documented on X posts [17]. Previous studies using location information in X have been conducted, such as to predict the prevalence of ZIKV cases in the United States by filtering tweets [18]. The studies used a bounding box as the boundary of the observed area, excluding tweets from Canada and Mexico. Spatial mapping shows similar patterns observed across the United States. Another study also uses geotagged tweet data, covering discussions about the #MeToo movement in the United States [19].

The results show further discourse about the movement, including the descriptions of abuse often related to the case. These existing studies suggested that geotagged tweets might provide information about ongoing social media discussions in a certain area.

This study aims to compare the discussion about masking in COVID-19 between the United States and Japan on social media and its changes throughout the pandemic stages. In the COVID-19 period, the 2 countries have different approaches to the response to the pandemic, resulting in different policies and rules. This difference was also reflected in how people responded, which can also be seen on social media. Mask-wearing, as one of the policies during COVID-19, generates a lot of discussion among the public. Some are supporting, and some are against mask-wearing. Despite their small number, people who did not want to wear masks are a vocal minority in sharing their opinion [20]. Due to all those differing opinions, while mask-wearing is indeed an important effort to mitigate the spread of COVID-19, there are still many debates about the necessity of masks [14,21]. A geographical focus in the analysis was also interesting to explore, as a survey showed that mask-wearing as a public health behavior might exhibit spatial heterogeneity, suggesting a different approach in the area based on country-level analysis in the United States [22].

In this study, we discussed the public opinions surrounding the COVID-19 mask mandate through geo-tagged tweets from the United States and Japan during the period from 2020 to 2022 in 3 stages of analysis: starting from its relevance to COVID-19 masks, stance for or against masking, and whether the tweets indicate users wearing masks. By discussing the factors and correlation of the observed tweets, we aim to understand individuals' stance towards the mask mandate and their actual mask-wearing behavior.

## Methods

### *Ethical Considerations*

This study did not require participants to be involved in any physical or mental intervention. This research also did not use personally identifiable information, thus exempted from institutional review board approval in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects stipulated by the Japanese national government.

### *Data*

The data in this study consists of geo-tagged tweets collected between January 1, 2020, and December 31, 2022, using Academic Research Access (discontinued in mid-2023). This dataset comprised 1,102,876 English tweets and 560,873 Japanese tweets. Note that we used the same dataset as our previous study on multilanguage data annotation [23]. Examples of paraphrased tweets showing their opinion about mask-wearing are shown in [Textbox 1](#).

**Textbox 1.** Example of tweets in English and Japanese.

## English:

- Against: Wearing this mask at work is exhausting. I'm sweating like crazy even from the smallest tasks.
- Not against: Unexpected perk of wearing a mask: it keeps your face nice and warm in freezing weather! \#MaskUp \#StayWarm \#WinterHack

## Japanese:

- Against: 普段あまりこういうことは言わないようにしてるけど、どうしてまだみんな外でマスクをしてるの？室内でも特に必要とは思えないし、もう外してもいいんじゃないかな、本当に。(I usually try not to say things like this, but why is everyone still wearing masks outside? I don't really see the need for them indoors either. Isn't it okay to stop wearing them already? Honestly.)
- Not against: それに、マスクをしていない人の割合もかなり高め。ちょっと気が緩みすぎてない？(Besides, the number of people not wearing masks is pretty high. Aren't people getting a little too careless?)

Demographic data were accessed from the Portal Site of Official Statistics of Japan [24] and Census Bureau data of the United States [25], obtained from the 2020 census. Previous studies have indicated that demographic characteristics such as sex, ethnicity, age, marital status, and employment status affect participation in preventive behaviors in the United

States [26]. In this study, we focused on several demographics for population characteristics at the state/prefecture level: total population, population by age, education level, information on foreign members/race, and income (Table 1). For Japan, the education level was obtained from the number of employed persons aged 15 years and older.

**Table 1.** Demographics variable used in correlation calculation.

Demographics	Variable
Population <sup>a</sup>	<ul style="list-style-type: none"> <li>• Total population</li> <li>• Population of working age<sup>b</sup></li> <li>• Population of elderly (older than 65 years)</li> </ul>
Education level <sup>c</sup>	<ul style="list-style-type: none"> <li>• Population completed bachelor or higher</li> </ul>
Income	<ul style="list-style-type: none"> <li>• Household income</li> </ul>
Race information	<ul style="list-style-type: none"> <li>• Race index (the United States), Number of foreign residents (Japan)</li> </ul>

<sup>a</sup>Total population from 2020 census.

<sup>b</sup>Percentage of people aged 18-64 years (the United States) and 15-64 years (Japan).

<sup>c</sup>Percentage of people completed bachelor or higher from population aged older than 25 years (the United States) from American Community Survey (2021) and population of employed person aged older than 15 years (Japan) from 2020 census.

## Preprocessing of Tweets

The first preprocessing step involved validating each tweet's location tag. Given the various types of location information in geo-tagged tweets, we focused on city-level information found in the "full-name" entity of the tweets. Instances where the geo-tagged location did not match the city lists in the United States and Japan were removed [27,28]. Next, we applied basic preprocessing steps, including changing usernames to common handles, addressing links, converting emojis to text, and removing duplicates and NAs.

Finally, we ensured that the tweets contained the keyword "mask" for English tweets and "マスク" (mask in Japanese) for Japanese tweets. In some cases, these keywords appeared as usernames or links, providing insufficient information or relevance to the topics and were thus eliminated in the process.

## Data Annotation

The annotation for our 1100 pairs of tweets was done by 4 annotators, 2 for each English and Japanese. We used Cohen kappa as the measure to evaluate interannotator agreement between annotators [29] and interpret the agreement results

[30]. The complete detailed approach and evaluation are explained in Ferawati et al [23]. To combine the annotated label between the annotators, we assigned the agreed label as the final label and assigned the disagreed label as not relevant, not against, and unknown for each stage.

## Classifier Training

X studies often use human annotators with the aid of machine learning classifiers [31,32]. We designed a multilingual guideline to obtain a suitable data annotation for this study [23]. The results of the annotation show a case of imbalanced data, with one category having a much smaller number of tweets compared with another. However, due to the time and resources required for human annotation, it is not practical to annotate a large number of tweets. To address this issue, we propose a semisupervised approach by using the BERT (Bidirectional Encoder Representations from Transformers) score from BERT-based classification method as additional data for training. By including additional training sets from predicted tweets, we hope to improve the performance of the classifier to solve the cases, especially in public perception and attitudes for COVID-19 masking in the 2 countries.

An existing study explained that BERT surpassed traditional machine learning algorithms, such as Logistic Regression, Support Vector Machine, Random Forest, and Naive Bayes, where BERT was supported by its ability to understand contextual information appearing in social media [33]. BERT was also widely used to analyze and derive conclusions. Based on the results of these studies, we decided to use BERT-based models for our study. We trained the annotated datasets using BERT-based classifiers, RoBERTa base model [34] for English tweets and BERT base Japanese [35] for Japanese tweets. To evaluate the models, we set aside 15% of the annotated data as the test set and applied 5-fold cross-validation to the remaining data. The final model was selected from the fold with the best results on the test set, evaluated using  $F_1$ -macro to address data imbalance.

Due to the limited number of annotated data, we included all 1100 annotated tweets. For each language, tweets annotated with the same categories by both annotators were assigned to those categories. In cases of conflicting annotations—for example, where annotator 1 marked a tweet as not relevant while annotator 2 marked it as relevant—we categorized the tweet as not relevant. The same approach was applied for the stance stage (categorized as unclear, then assigned to not against) and mask-wearing (categorized as unknown).

### Semisupervised Approach

We observed an imbalance in the categories of our annotated data, which likely contributed to the low performance of our models. Due to constraints in obtaining additional annotated data, we implemented a semisupervised approach for the stance and mask-wearing stages of the tweets. The semisupervised approach was a common measure taken to obtain more annotated data without a significant increase in annotation costs [36]. In this study, we stored the BERT outputs from the predictions and filtered them based on determined thresholds for each category. We considered 2 methods of adding pseudo-labeled data: adding an equal number of tweets to both categories and adding tweets only to the minority category. Using softmax-transformed values of the BERT output for each class, we selected a set of data to be added to the training. The objective for implementing the

approach is to obtain labels with a higher confidence from the obtained softmax-transformer values as an additional sample for training. We maintained the chosen split of training and validation from the previous step in classifier training.

The selected tweets from this step were then included as additional training data using the same parameters and model as in the previous subsection. The data for validation and testing remained the same as the baseline. We experimented with several combinations of thresholds for each class in each stage (against and not for stance; no, yes, and unknown for mask-wearing), while endeavoring to keep the number of selected additional tweets balanced. When determining the threshold value for the selection, we selected a certain value to balance our objectives: retaining sufficient confident predictions for the reliability of our pseudo-label and having enough additional samples to meaningfully expand our dataset. The threshold was set by observing the transformed BERTScore from the prediction results, plotting them into a histogram with the aid of descriptive statistics of the score to determine the appropriate threshold for filtering. We then compared the results of each training.

### Correlation

We calculated the correlation between the predicted percentages and demographic data from the 2 countries, the United States and Japan, using Pearson correlation. In this study, we considered 3 levels of significance for the  $P$  values with a significance level of .05, .01, and .001.

## Results

### Interannotator Agreement

For the annotation phase, we used only a subset of the entire dataset. Considering the reasonable workload of our annotators, we selected 1100 tweets from each language, as reported in another publication focusing on the annotation of a multicultural dataset [23]. Although the agreement levels for stance and mask-wearing stage are still less than  $<0.60$ —a score that was categorized as moderate by Cohen (1960) [29] but inadequate by McHugh (2012) [30]—they are deemed acceptable for our study (Table 2).

**Table 2.** Cohen kappa for the annotation results. The percentage in each round shows the number of samples annotated. The agreement is calculated for each round and each part of the samples.

Stage	Language	
	English	Japanese
1: Relevancy	0.79	0.92
2a: Stance	0.46	0.59
2b: Mask-wearing	0.55	0.47

The number of tweets varies substantially across the country, with certain states or prefectures having far more tweets than others. The difference in number is primarily observed in densely populated areas, such as the capital or popular destinations. Among tweets identified as relevant to

COVID-19, California leads the United States in mask-related discussions, while Tokyo tops the list in Japan. A list showing the top and bottom tweet counts for states (in the United States) and prefectures (in Japan) is provided (Table 3).

**Table 3.** Top and bottom 3 of tweet counts in the United States and Japan.

Ranking	United States		Japan	
	State	Tweet count	Prefecture	Tweet count
Top 3	California	113,148	Tokyo	85,836
	Texas	75,517	Kanagawa	27,737
	New York	60,317	Osaka	26,725
Bottom 3	Vermont	1314	Kochi	1197
	North Dakota	1079	Tottori	1149
	Wyoming	776	Shimane	1049

## Classifier Training

We fine-tuned roberta-base and bert-base-japanese-v3 with our annotated tweets to predict the rest of the data, a method commonly used for social media data [37,38]. We trained the classifier using this annotated data and then used it to predict the remaining tweets. The parameters for our final model used the AdamW optimizer, with a batch size of 32, 10 epochs with early stopping callback, learning rate of 0.00002, and weight decay of 0.01.

We used the entire sample of 1100 tweets for relevancy classification (Table 4). In the first stage of relevancy classification, we achieved the highest  $F_1$ -macro scores for both English and Japanese texts compared with the majority baseline. However, the stance and mask-wearing stage did not yield similar results, showing only a marginal increase from the majority baseline. This indicates that the models are still unable to correctly predict the minority categories. These results also mirror the annotation results, where annotators struggled to achieve higher interannotator agreement.

**Table 4.** Results of BERT (Bidirectional Encoder Representations from Transformers)-based classifier. Only tweets annotated as relevant by both annotators are included as a sample for Stage 2a: Stance and Stage 2b: Mask-wearing. As the category is heavily imbalanced for all the stages, we calculated the majority baseline by assigning majority classes as the predicted category in each stage: relevant for relevancy, not against for stance, and unknown for mask-wearing.

Stage	Language	Number of samples, n	Majority baseline	$F_1$ -macro
1: Relevancy	English	1100	0.44	0.78
	Japanese	1100	0.39	0.75
2a: Stance	English	880	0.47	0.61
	Japanese	706	0.45	0.69
2b: Mask-wearing	English	880	0.31	0.31
	Japanese	706	0.29	0.40

## Semisupervised Approach and the Final Classifier

By adopting a semisupervised approach to train our classifier, we were able to improve second-stage  $F_1$ -scores that were nearly as high as those from the first stage—0.68 for English tweets and 0.72 for Japanese tweets.

Based on the results for the stance stage, the chosen threshold with a more balanced number of tweets in each category (eg, an additional 650 in the “against” category, bringing the total to 755 “against” and 775 “not against” for English data) achieved a better  $F_1$ -macro score than other alternatives.

When it comes to classifying mask-wearing behavior, maintaining a balanced sample for re-training is challenging due to the imbalance of labels in the annotated data. In English tweets, only 46 tweets were predicted as “not wearing mask” out of the entire dataset, so we included all the

predicted tweets in the model. The best result was an  $F_1$ -score of 0.487, a substantial increase from the previous score of 0.301 with the original annotated data. For Japanese tweets, our best model only yielded an  $F_1$ -score of 0.48.

For the mask-wearing stage, the distribution of BERT score outputs was much more skewed than for stance in both languages. Because it was challenging to maintain balanced numbers for each category, we experimented with various threshold combinations to assess performance. Due to the difficulty in balancing the numbers, the improvement in the model performance seems rather limited. Therefore, it should be considered the best our classifiers could achieve and is acceptable to be used in our following tasks.

The final classifier model was trained on a combined dataset of the original annotated data and the semisupervised data. The remaining tweets were classified using the final model. We deemed the  $F_1$  -score from the previous semisupervised results acceptable for our study and used it as a

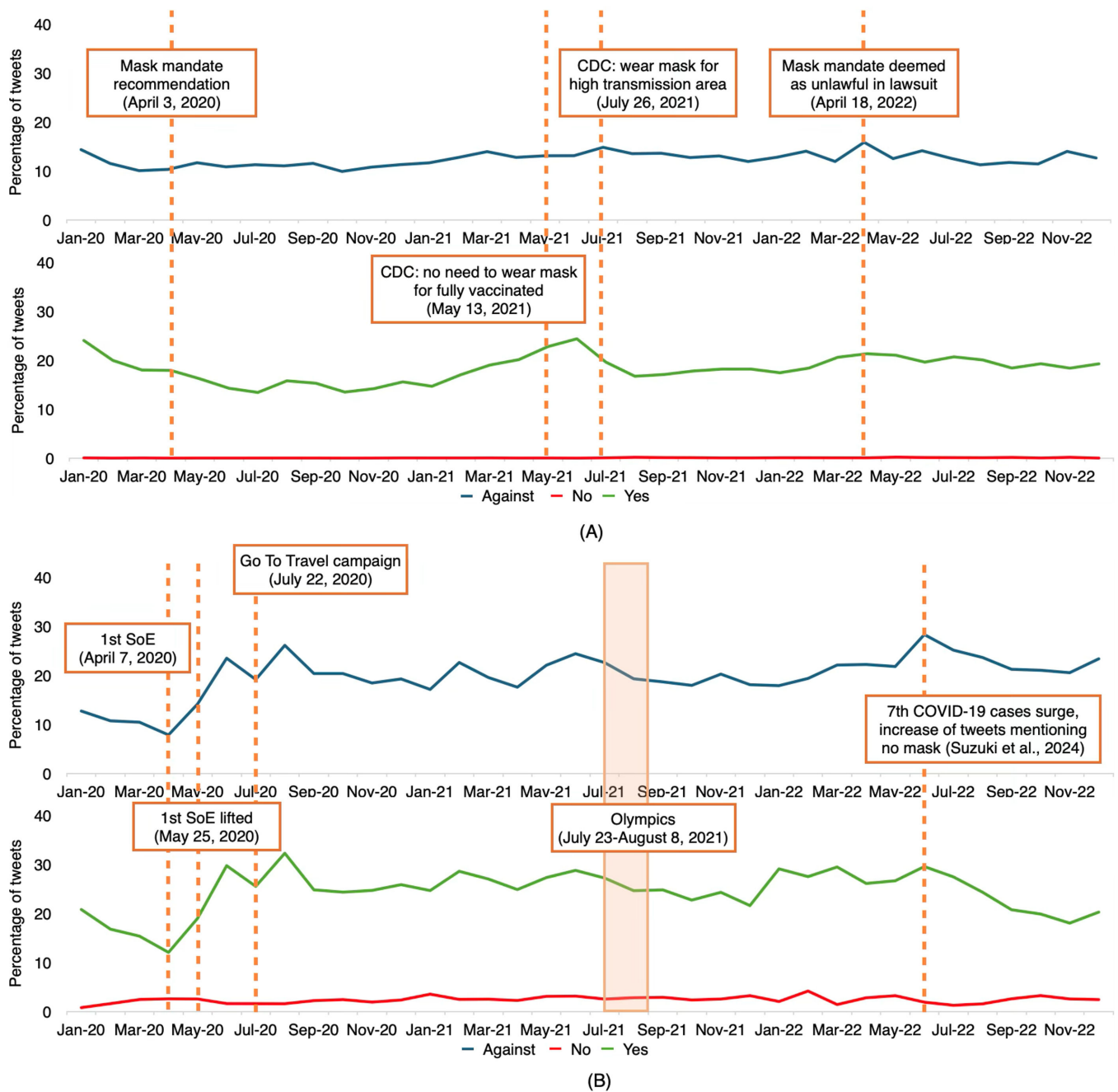
basis for generating a map of stance toward the mask mandate and actual mask-wearing behavior from related tweets in the United States and Japan, using both English and Japanese tweets.

### Time Series Observation of the Classification Results

The tweets were collected over a 3-year period, capturing the trend from early, peak, to post-peak periods of the COVID-19

pandemic. Throughout this time, various events and policies occurred, potentially influencing tweet volume and content (Figure 1). In the figure, no observable trend in the stance against COVID-19 masking is evident in the United States, suggesting little change in people’s opinion about mask-wearing from early 2020 to the end of 2022. Both countries show a decreasing trend in tweets against masking from January 2020 to April, followed by an upward trend after relevant departments released recommendations about mask-wearing.

**Figure 1.** Time series plot of tweets against masking in (A) the United States and (B) Japan. Some events and policies happening throughout the period are highlighted in the figure. The upper part of the graph shows the percentage of tweets against mask-wearing, while the lower part shows tweets mentioning users wearing a mask (green) and not wearing a mask (red) [39]. CDC: Centers for Disease Control and Prevention; SoE: state of emergency.



Throughout the observed period, the CDC released several guidelines on the utilization of masks as a preventive measure. They initially mandated the use of masks in April 2020 and later revised the recommendation after the development of vaccines, such as lifting the recommendation to wear masks for fully vaccinated individuals in May 2021. In the month following CDC's revision of their mask-wearing guidance to no need to wear a mask if vaccinated, tweets classified as "not wearing mask" reached their lowest point, while tweets classified as "wearing mask" peaked (Figure 1). Subsequently, the CDC released another recommendation to wear masks in high-transmission areas in July 2021, which is reflected in interesting changes observed in the time series plot of mask-wearing stages. The number of tweets indicating "not wearing mask" substantially increased from 0.03% in June 2021 to 0.21% in August 2021, while tweets about "wearing mask" saw a decrease from 24.41% in June 2021 to 16.76% in August 2021. In April 2022, as a result of a lawsuit regarding mask mandates deemed unlawful in Florida, nationwide mask mandates for airplanes and public transportation were no longer enforced [40]. As shown in Figure 1A, in the same month, a peak was observed in the stance against mask mandates in the United States, and tweets classified as "not wearing mask" peaked a month after the ruling, observed in May 2022.

In Japan, there are 3 noticeable fluctuations in the figure, closely related to COVID-19 policies. We observed some significant changes when the government announced a state of emergency (SoE) in April 2020 and subsequently lifted it in May 2020. After a dip in the percentage of tweets expressing a stance against masking after the first SoE in April 2020, the first peak appeared a month after the first SoE was lifted. These fluctuations are also consistent with findings from another study on responses to COVID-19 waves in Japan [41]. Another study discussed that there are indeed signs of disruptions due to COVID-19 in X,

which shows top concerns during the SoE period and some impacts of the pandemic on societies in Japan [42]. The second peak appeared at the beginning of the Olympics. The Tokyo Olympics, held from July 23 to August 8, 2021, also influenced the observed changes in the figure. Before the Olympics began, in June 2021, there was a slight increase in tweets expressing a stance against mask-wearing, as shown in Figure 1B. The third peak appeared in June 2022, coinciding with the 7th surge in COVID-19 cases and an increase in "no mask" tweets, as observed by Suzuki et al [39].

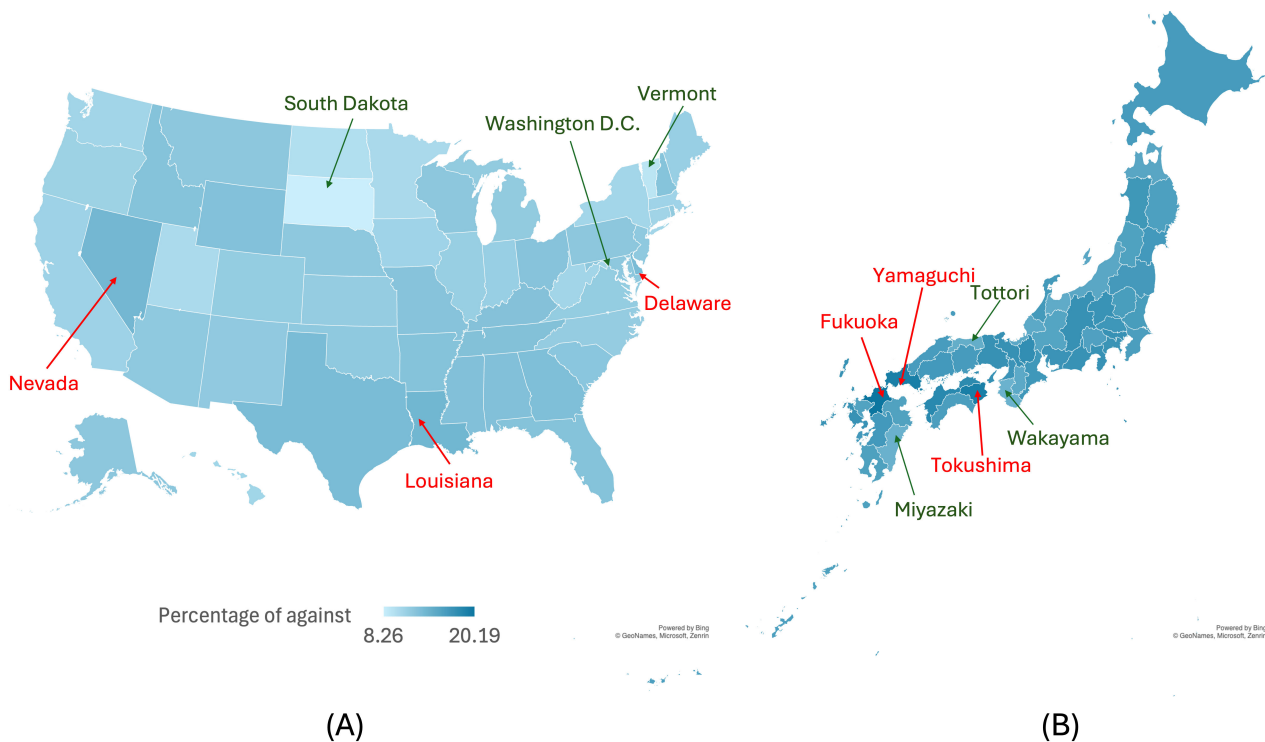
## Map Comparison

### Tweets Against Masking

The final classifications from the selected model were presented in the form of maps, depicting the percentage of tweets classified as against mask-wearing in the United States (at the state level) and Japan (at the prefecture level). Table 3 reveals a significant discrepancy in tweet frequencies between states and prefectures. For instance, California has over 100,000 relevant tweets, while other states like Wyoming have fewer than a thousand tweets. Similarly, in Japan, Tokyo, the capital, has the highest number of relevant tweets, with over 85,000, while Tottori has the fewest tweets, with approximately 1000.

In the United States, Louisiana, Nevada, and Delaware were the 3 states with the highest percentage of tweets against masking, ranging from 13.28% to 13.75% (Figure 2). The 3 states with the lowest percentage are Washington, DC; Vermont; and South Dakota, with about 8%-9%. On the other hand, in Japan, the percentage of tweets against masking is observed to be higher, ranging from 19.28% to 20.19% in Fukuoka, Yamaguchi, and Tokushima prefectures, with the lowest percentage observed at about 14% in Tottori, Miyazaki, and Wakayama prefectures.

**Figure 2.** Percentage of tweets against masking in (A) the United States and (B) Japan. The same color scale is used for both the United States and Japan. The top 3 prefectures/states with the highest percentage are marked in red text and arrow, while the bottom 3 are marked in green text and arrow.



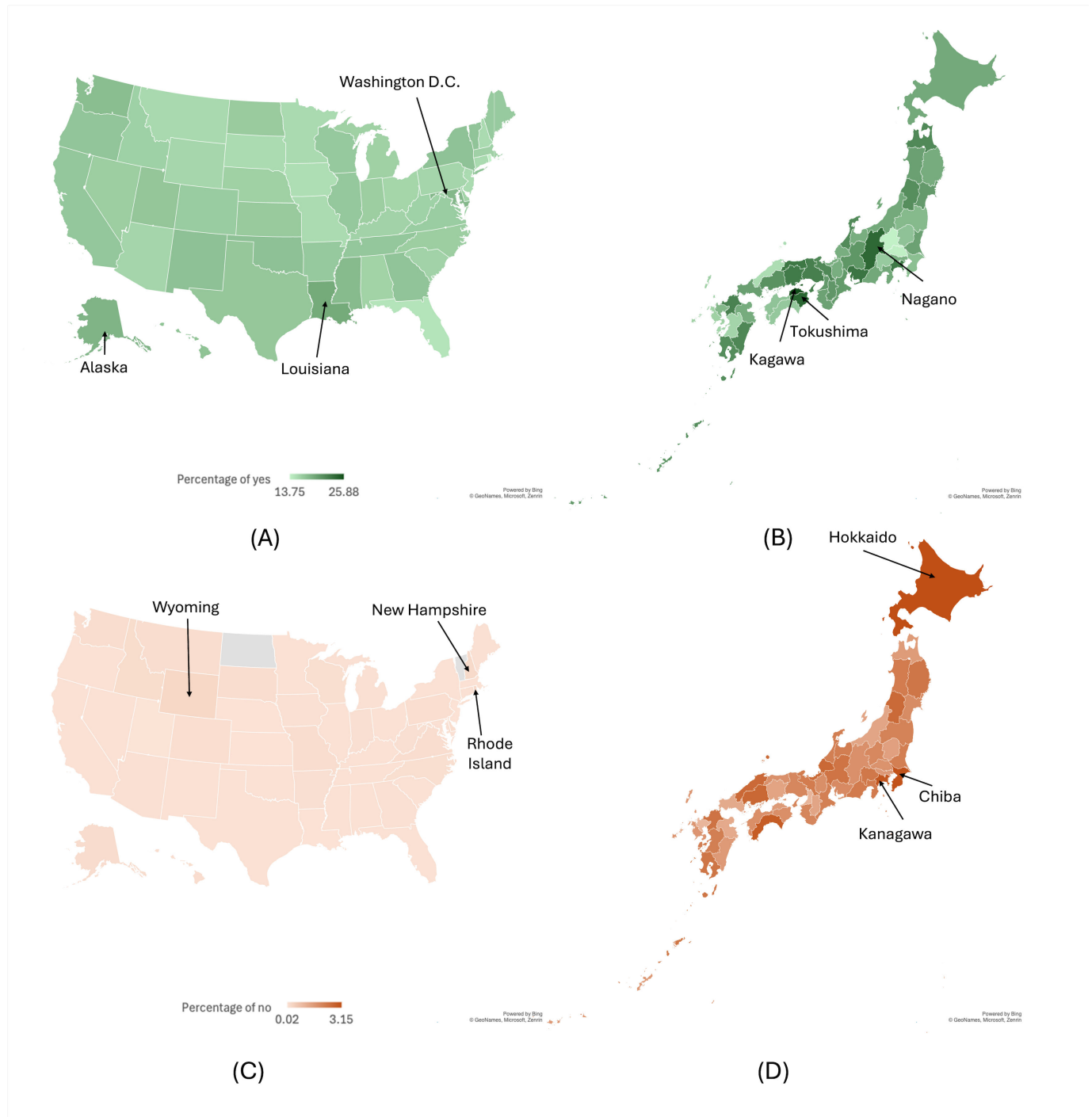
### Mask-Wearing Classification Results

The analysis and maps for mask-wearing behavior tweets were divided into 2 categories: those classified as “not wearing” and “wearing” masks (Figure 3). Since some states or prefectures have a much smaller number of tweets, the count of tweets classified as “not wearing” is reduced accordingly. For English tweets, the total count of “not wearing” tweets was very small, at 574, resulting in no tweets classified in the category in several states and causing the percentage to appear as 0. Wyoming has the highest percentage of tweets classified as “not wearing,” at 0.26% (2 out of 776). In contrast, Louisiana leads in the percentage of “wearing” mask tweets, at 19.12% (2122 out of 11,100); followed by Washington, DC (1638 out of 8845, 18.52%); and Alaska with about 18.33% (333 out of 1817). However,

a careful interpretation is necessary for these figures, as the number of tweets was relatively small, especially for the tweets mentioning not wearing masks. The tweets were much smaller in the behavioral aspects of mask-wearing, so the resulting maps might not have enough data for a state-level conclusion.

On the Japan side, there were 8462 tweets classified as “not wearing” masks, a higher number compared with the United States results. However, a big portion of these tweets was still concentrated in Tokyo, with over 2000 tweets, while the other 31 prefectures had fewer than 100 tweets each. In terms of the percentage of tweets classified as “not wearing” masks, Hokkaido has the highest percentage at 3.15%. For tweets classified as “wearing” masks, Kagawa has the highest percentage at 25.87%.

**Figure 3.** Percentage of tweets classified as wearing a mask in (A) the United States and (B) Japan, and tweets classified as not wearing a mask in (C) the United States and (D) Japan. The state/prefecture name shown denotes the top 3 areas with the highest percentage. The same color scale is used for both the United States and Japan. Gray shows no predicted data for the area.



**Correlation With Demographics Data**

We calculated the correlation between stance percentages and the demographic data from the country census for both the United States and Japan (Table 5). In the United States, we observed a negative correlation of  $-0.384$  between household income and stance against masking percentages. Although this correlation is relatively low, it is still significant within the 1% level of significance threshold. This suggests that states with higher incomes tend to have lower percentages of tweets against masking. Additionally, education level also

shows a significant negative correlation, showing that with an increase in the population completing higher studies, there was an observed decrease in tweets with a negative stance towards the mask mandate. In Japan, we found that the number of people older than 65 years showed a positive correlation with the percentage of tweets against the mask mandate at the prefecture level. Similar results were observed in the number of people who completed a bachelor or higher degree.

**Table 5.** Correlation calculation of tweets percentage with the demographics data from the United States and Japan. The table only displayed a significant correlation.

Variable	Correlation	95% CI	P values
The United States			
Population completed bachelor and higher	-0.456	-0.653 to -0.200	.001
Household income	-0.353	-0.577 to -0.080	.013
Japan			
Population of elderly (older than 65 years)	0.295	0.009 to 0.537	.044
Population completed bachelor and higher	0.310	0.024 to 0.548	.034

## Discussion

### Principal Results

#### Key Events That Drove the Surge in COVID-19-Related Tweets

The time series plot illustrates the changes in stance against masking and mask-wearing tweets in the United States and Japan. The major fluctuations in the data points often coincide with governments announcing mask-related policies or major public events such as lockdown or the Olympics during the COVID-19 period. Such a phenomenon suggested that individuals tend to be prompted to express themselves on social media by major movements from government or public events, as previous studies also indicated [43-46]. In addition to the temporal trend, the geographical visualizations from the final results provide a general landscape of public responses to the mask-related policies divided by the municipal regions on social media (referred to as “X” in this study) in the United States and Japan. Information regarding the percentage of tweets against masking is displayed for each state or prefecture (Figure 3). Interestingly, Washington, DC, emerges as the state with the lowest proportion of tweets against masking and the highest percentage of tweets mentioning mask-wearing. This suggests that the state exhibits less resistance to mask-wearing and actively follows mask guidance.

Based on the correlation results between tweets and demographic data, education level and household income were found to be significantly correlated in the United States. Specifically, the population of people who have completed a bachelor’s degree and higher is negatively correlated with a stance against masking. This result indicates that with the increase in the higher education levels of the population in the states, we observed a decrease in the number of tweets against masking. This observation aligns with findings from a previous study based on a survey in the United States, which reported that individuals with at least a college degree are less likely to refrain from wearing face masks [47]. Additionally, the same study found that households with lower incomes were more likely to exhibit the same behavior of not wearing masks, an insight consistent with our findings.

Previous studies indicate that the behavior of older and younger people changes throughout the pandemic, with older individuals becoming more cautious about preventive measures as the pandemic progresses [48]. Although the

majority of X users fall within the young adult category (aged 18-29 years), there are still a considerable number of older X users in the United States, as revealed by a survey conducted in 2023 [49]. Since this study calculated correlations for the entire period of data collection, these changes in opinion might not have been captured well. Throughout the study period, there have been several changes in mask-wearing rules, compliance, and when masks are mandated in some areas, which also affects how people responded and their opinion towards mask-wearing [50]. New information about mask-wearing released by the authority can also result in changes in public attitude toward mask-wearing [51], such as when people initially support mask-wearing and then change their opinion, or if they are against masking in the beginning and change to accepting masks in the later stage of the pandemic. This was also supported by results of the study, where a change in the sentiment is observed in the 3 major cities in the United States [15].

In Japan, the number of people aged older than 65 years shows a positive correlation with the number of tweets expressing their stance against masking, in line with a previous study in Hiroshima, Japan [52]. Based on existing studies summarizing responses in 27 countries, including Japan, people over 60 did not really comply with wearing masks outside of their home [53], which is also reflected in the results of this study [54]. The overall results of the correlation calculated are, however, quite small. Such a result indicated that there could be more confounding factors influencing the results. Below, we discuss some potential factors suggested in previous literature.

Health policy differences in the states within the United States might be one of the possible factors affecting the mask-wearing behavior reflected in social media. There had been observable differences, such as physical distancing rules across states and the relation of the measure to racial and socioeconomic differences in the area [55]. There was also an existing study documenting mask-wearing policy across the states and its enforcement [56]. The result indicates that overall, the mask is not strongly enforced in the states. The resistance to mask-wearing in the United States could also be viewed through cultural aspects of people living in the country [57]. A study focusing on the cultural and political aspects of mask-wearing suggested that future attempts at mask-wearing policy need to consider the cultural meaning of masks to the public [58].

A difference in response to policies was observed in Japan, with results from a study showing the high trust and compliance in government policies during COVID-19 [59]. An internet survey showed that some habits of the Japanese, such as wearing masks prior to the pandemic, were also related to their responses to the preventive measures for COVID-19 [60]. These policies and behavioral insights might serve as additional variables to further explain the observed web-based behavior in X. The insights can be used in handling information dissemination in social media, for example, by the government in countries to design the most effective way to communicate with the public on social media.

### Interpreting Conflicting Correlations: Mobility, Social Media, and Educational Shifts

Most of our findings aligned with previous research. However, unexpectedly, individuals with a bachelor's degree or higher also showed a positive correlation with opposition to the mask mandate, which contradicts earlier studies [48]. One possible cause for the inconsistency was the difference in the demographics between people using X and the census data used for the correlation calculation. In the United States, the percentage of X users with at least a bachelor's degree is higher than that of the general public, which is also observed in the income levels [61]. In Japan, based on a survey conducted by the Ministry of Internal Affairs and Communications, about 80% of people in their 20s were X users. The penetration rate is the highest in the younger population, but there were also users in their 60s (18.9%) [62]. According to a study conducted using X data, even within the social media users' demographics, there was a bias due to the choice to participate in a certain topic for social media discussion [63]. This disparity might affect the correlation results with tweet percentages, as the demographics being calculated are not X-specific demographics but public demographics based on census and surveys.

Another possible cause of the inconsistency in results may also have been caused by the mobility of people throughout the COVID-19 period in Japan. The tweet data were collected over a 3-year period, while the education level was obtained from the 2020 census. A previous study mentioned that before the pandemic, people were centralized in the Tokyo metropolitan area, a pattern disrupted during COVID-19 [64]. The changes in migration rates were also influenced by the emergence of remote work during the pandemic, prompting younger people to move to rural areas.

The increase in tweets mentioning “no-mask,” as shown in the study by Suzuki et al [39], might also contribute to the conflicting correlation results, particularly in the latter half of 2022. Another reason might be caused by changes in the education style during the early period of COVID-19 to adapt to the changing policy. The changes to web-based education proved to be overwhelming for students, with their self-reported mental health status being worsened [65]. The anxiety and depression experienced during COVID-19 also increased students' social media usage [66]. This might be reflected in opinions about COVID-19 and its measures,

such as mask-wearing, where similar cases were observed in studies about disruptions caused by COVID-19 in social media [42].

Our findings indicated that mask-wearing behavior was observed in social media discourse in both the United States and Japan. The map showing the area with a high percentage of yes and no to mask-wearing can be considered a guideline in how to communicate or release guidelines in the event of a pandemic, with more attention to the area with a high percentage of “no” in mask-wearing. Based on the number of tweets, we may also launch more information on social media to reach more people in the area. The correlation results show some significant correlations observed with population data of the states and prefectures, with some of them consistent with existing findings. These results suggest that the social media data has the potential to aid in research, with some caution due to the limited nature of the geo-tagged data in social media. Focusing on other aspects of analysis, such as other potential factors in correlation analysis or additional data, can be a future research direction to improve the results.

The Health Belief Model was believed to be useful for understanding and investigating the participation in COVID-19 preventive measures in the United States [67]. A study using the Understanding America Survey showed that the Health Belief Model results were consistent with existing studies, with mask-wearing behavior positively endorsed as one of the preventive measures. A study comparing public perceptions in different countries during the pandemic, Japan and the United States included, also showed that higher perceived effectiveness was one of the factors affecting how the public responded to preventive behavior, such as mask-wearing or handwashing [68]. To make the public more accepting of preventive measures or policies, the public needs to trust the effectiveness of the act. The role of publishing and spreading the effectiveness can be entrusted to social media, with emphasis in areas with higher percentages in the stance of not wearing masks.

### Limitations

This study focuses solely on the English and Japanese languages for analyzing short and informal text, specifically tweet posts. In the second stage, we encountered numerous ambiguous and unclear tweets regarding attitude. Analyzing such stages proved challenging due to the short text and ambiguity of tweet content, which was often insufficient for accurately determining attitudes. Therefore, we chose to exclude a neutral option, limiting choices to either against or not.

Although X data might be subject to sampling bias due to its limited user base, the data were proven useful for infodemiology research, especially in identifying and monitoring public discussion topics during the pandemic [69,70]. Population-wise, our sample is restricted to individuals who voluntarily provide location information in their tweets. Despite the limitation of such data, it reflects the opinion of the users to a degree, as reported in an existing study that showed that geotagged tweets were able to display public opinion of people in the United States toward

COVID-19 vaccinations [71]. Disparities in tweet counts across states/prefectures may influence the final percentage results displayed in the figures, which are based on the entire 3 years of data. We calculated the percentage based on the tweets observed in each state as one of the measures to address the potential problems due to the location data. The number of tweets in each state varies, so we avoid general bias caused by comparing the tweets based on their count, but based on the percentage of the classified tweets. Due to the data collection period and collective analysis after, the resulting figure, especially in the form of map visualization, may not sufficiently capture early COVID-19 period changes and post-pandemic trends shown in the time series plot.

A previous study identified around 8.4% of tweets as bots from around 200,000 accounts collected in the early period of the pandemic, March to June 2020 [72]. Another study mentioned that bots focused more on political content in the COVID-19 pandemic discourse compared with health content, such as actions taken by the public during the pandemic [73]. The bots mostly discussed information related to COVID-19 or pandemic statistics. In this study, we focused on mitigating bots' influence, such as artificial amplification by removing duplicate tweets, one of which often appeared in posts by bot accounts. This action ensured that no single automated message could statistically dominate our social media dataset or skew our demographic correlations. To verify this approach, we conducted a manual audit in the annotated dataset, which was sampled randomly from the entire tweets data. We did not find apparent bot accounts during the check in our sample. Given this minimal presence, we conclude that the presence of bots does not significantly skew the overall stance analysis or correlations in our study.

The limited number of "not wearing" mask annotations constrained the number of predicted instances in the rest of the tweets. Consequently, the final predicted percentages for "not wearing" masks were also very small, only 0.07% for English tweets and 2.33% for Japanese tweets. This problem, caused by a small number of tweets annotated as "not wearing mask" was also further added by a surge in tweets mentioning "no-mask" after June 2022 [39]. This surge, as documented by Suzuki et al [39], may contribute to the disparity from previous studies and introduce bias into the correlation calculations, requiring people to interpret the results cautiously. To address these challenges, future improvements in classification and additional annotations are necessary for enhanced accuracy.

The initial score of the classification results obtained was not satisfactory. The  $F_1$ -score for stance against masking is 0.61 for English and 0.69 for Japanese. It was even smaller for the mask-wearing stage, with 0.31 for English and 0.40 for Japanese. While it is higher than the majority baseline, it

is still not performing as well for the classification. There is a slight increase in the  $F_1$ -score after adding pseudo-labeled data, as observed in the stance and mask-wearing stage. However, the scores were still not as high, indicating that the topic to be classified is indeed challenging, as both human annotators and classifiers struggle to achieve a high score in classifying stance against masking and mask-wearing information of the users.

Regarding the mask-wearing stage, the number of tweets annotated as "not-wearing" masks is very small in the annotation phase (10 for English and 11 for Japanese). This small number affects the classification because when we split the data, we were working with an even smaller number of tweets for training, validation, and testing. This makes it even harder to classify the "not-wearing-mask" tweets, resulting in a low  $F_1$ -score in the testing phase. This issue is further highlighted in the classification of the remaining tweets, with only 0.07% predicted as "not-wearing" masks in English and 2.33% in Japanese out of all the predicted tweets. The low number of tweets predicted as "not-wearing" masks is also apparent in Figure 3, especially in the United States case. This limitation warrants caution in interpreting our results, especially since the annotation target was particularly challenging.

Due to the policy of X, we are unable to obtain the demographics of its users; thus, we considered using the demographics of the general population in the country. However, there are differences between the demographics of X users and the general population [61,62] which might affect the results of the correlation calculated in this study.

## Conclusions

Our findings revealed a clear shift in public opinions at several key moments, often coinciding with major policy changes or public events, such as mask mandates and the Olympics, observed from 1,102,876 English tweets and 560,873 Japanese tweets from the United States and Japan between early 2020 and late 2022, covering the entire duration of the COVID-19 pandemic. Based on the correlation observed in the social media data, some variables are significantly correlated with the number of tweets, such as education levels, the population of the elderly, and household income. Some of our findings align with previous research across major social media platforms, suggesting that public opinions reflected in social media data are often interconnected. However, divisions based on education level, household income, and generational differences also indicate that social media interactions may be inherently biased, partly due to recommendation algorithms that encourage clicks and foster echo chambers.

## Funding

This work was supported by JST SICORP (grant number JPMJSC2107), JSPS KAKENHI (grant number JP22K12041), and the Cross-ministerial Strategic Innovation Promotion Program (SIP) on "Integrated Health Care System" (grant number JPJ012425), Japan.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to X's data sharing policy, but the IDs are available from the corresponding author on reasonable request.

### Authors' Contributions

KF collected the data. KF, WJS, and SW conducted the experiments and analyzed the results. SW and EA provided critical comments. All authors contributed to the design and execution of the study and reviewed the manuscript. EA supervised the project.

### Conflicts of Interest

None declared.

### References

1. Hale T, Angrist N, Kira B, Petherick A, Phillips T, Webster S. Variation in government responses to COVID-19. Blavatnik School of Government. 2020. URL: <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker> [Accessed 2024-01-12]
2. Netburn D. A timeline of the CDC's advice on face masks. Los Angeles Times. 2021. URL: <https://www.latimes.com/science/story/2021-07-27/timeline-cdc-mask-guidance-during-covid-19-pandemic> [Accessed 2024-02-14]
3. Näher AF, Schulte-Althoff M, Kopka M, Balzer F, Pozo-Martin F. Effects of face mask mandates on COVID-19 transmission in 51 countries: retrospective event study. *JMIR Public Health Surveill.* Mar 8, 2024;10:e49307. [doi: [10.2196/49307](https://doi.org/10.2196/49307)] [Medline: [38457225](https://pubmed.ncbi.nlm.nih.gov/38457225/)]
4. Miura A, Murayama A, Kitamura H. Behind the mask: analyzing the dual influence of social norms on pandemic behavior in Japan 1. *Jpn Psychol Res.* 2024. [doi: [10.1111/jpr.12520](https://doi.org/10.1111/jpr.12520)]
5. Noguchi K. Examination of the content of individualism/collectivism scales in cultural comparisons of the USA and Japan. *Asian J Social Psychol.* Sep 2007;10(3):131-144. [doi: [10.1111/j.1467-839X.2007.00220.x](https://doi.org/10.1111/j.1467-839X.2007.00220.x)]
6. Ishii K, Eisen C. Socioeconomic status and cultural difference. *Oxf Res Encycl Psychol.* 2020. [doi: [10.1093/acrefore/9780190236557.013.584](https://doi.org/10.1093/acrefore/9780190236557.013.584)]
7. Uricher R, Nakayama M, Uchida Y. Cultural differences in behavioral regulation under the collective threat of COVID-19: more adjustment in Japan and more influence in the United States. *J Appl Social Psychol.* Jul 2025;55(7):483-496. [doi: [10.1111/jasp.13103](https://doi.org/10.1111/jasp.13103)]
8. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE.* Nov 29, 2010;5(11):e14118. [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
9. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE.* May 4, 2011;6(5):e19467. [doi: [10.1371/journal.pone.0019467](https://doi.org/10.1371/journal.pone.0019467)] [Medline: [21573238](https://pubmed.ncbi.nlm.nih.gov/21573238/)]
10. Dyer J, Kolic B. Public risk perception and emotion on Twitter during the Covid-19 pandemic. *Appl Netw Sci.* 2020;5(1):99. [doi: [10.1007/s41109-020-00334-7](https://doi.org/10.1007/s41109-020-00334-7)] [Medline: [33344760](https://pubmed.ncbi.nlm.nih.gov/33344760/)]
11. Su Y, Venkat A, Yadav Y, Puglisi LB, Fodeh SJ. Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities. *Comput Biol Med.* May 2021;132:104336. [doi: [10.1016/j.combiomed.2021.104336](https://doi.org/10.1016/j.combiomed.2021.104336)] [Medline: [33761419](https://pubmed.ncbi.nlm.nih.gov/33761419/)]
12. Imran M, Castillo C, Diaz F, Vieweg S. Processing social media messages in mass emergency. *ACM Comput Surv.* Jul 21, 2015;47(4):1-38. [doi: [10.1145/2771588](https://doi.org/10.1145/2771588)]
13. Thorpe Huerta D, Hawkins JB, Brownstein JS, Hswen Y. Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: a Twitter analysis. *SSM Popul Health.* Sep 2021;15:100851. [doi: [10.1016/j.ssmph.2021.100851](https://doi.org/10.1016/j.ssmph.2021.100851)] [Medline: [34355055](https://pubmed.ncbi.nlm.nih.gov/34355055/)]
14. Wang Y, Hao H, Platt LS. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter. *Comput Human Behav.* Jan 2021;114:106568. [doi: [10.1016/j.chb.2020.106568](https://doi.org/10.1016/j.chb.2020.106568)]
15. Saito R, Tsugawa S. Understanding citizens' response to social activities on Twitter in US metropolises during the COVID-19 recovery phase using a fine-tuned large language model: application of AI. *J Med Internet Res.* Feb 11, 2025;27:e63824. [doi: [10.2196/63824](https://doi.org/10.2196/63824)] [Medline: [39932775](https://pubmed.ncbi.nlm.nih.gov/39932775/)]
16. Deiner MS, Deiner NA, Hristidis V, et al. Use of large language models to assess the likelihood of epidemics from the content of tweets: infodemiology study. *J Med Internet Res.* Mar 1, 2024;26:e49139. [doi: [10.2196/49139](https://doi.org/10.2196/49139)] [Medline: [38427404](https://pubmed.ncbi.nlm.nih.gov/38427404/)]
17. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. *JMIR Public Health Surveill.* Apr 5, 2021;7(4):e26780. [doi: [10.2196/26780](https://doi.org/10.2196/26780)] [Medline: [33720841](https://pubmed.ncbi.nlm.nih.gov/33720841/)]
18. Masri S, Jia J, Li C, et al. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. *BMC Public Health.* Jun 14, 2019;19(1):761. [doi: [10.1186/s12889-019-7103-8](https://doi.org/10.1186/s12889-019-7103-8)] [Medline: [31200692](https://pubmed.ncbi.nlm.nih.gov/31200692/)]

19. Modrek S, Chakalov B. The #MeToo movement in the United States: text analysis of early Twitter conversations. *J Med Internet Res*. Sep 3, 2019;21(9):e13837. [doi: [10.2196/13837](https://doi.org/10.2196/13837)] [Medline: [31482849](https://pubmed.ncbi.nlm.nih.gov/31482849/)]
20. Taylor S, Asmundson GJG. Negative attitudes about facemasks during the COVID-19 pandemic: the dual importance of perceived ineffectiveness and psychological reactance. *PLoS ONE*. 2021;16(2):e0246317. [doi: [10.1371/journal.pone.0246317](https://doi.org/10.1371/journal.pone.0246317)] [Medline: [33596207](https://pubmed.ncbi.nlm.nih.gov/33596207/)]
21. Eikenberry SE, Mancuso M, Iboi E, et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model*. 2020;5:293-308. [doi: [10.1016/j.idm.2020.04.001](https://doi.org/10.1016/j.idm.2020.04.001)]
22. Taube JC, Susswein Z, Bansal S. Spatiotemporal trends in self-reported mask-wearing behavior in the United States: analysis of a large cross-sectional survey. *JMIR Public Health Surveill*. Mar 6, 2023;9(1):e42128. [doi: [10.2196/42128](https://doi.org/10.2196/42128)] [Medline: [36877548](https://pubmed.ncbi.nlm.nih.gov/36877548/)]
23. Ferawati K, She WJ, Wakamiya S, Aramaki E. Synchronizing approach in designing annotation guidelines for multilingual datasets: a COVID-19 case study using english and japanese tweets. In: Prabhakaran V, Dev S, Benotti L, Hershovich D, Cabello L, Cao Y, Adebara I, Zhou L, editors. Presented at: Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP; Aug 16, 2024:32-41; Bangkok, Thailand. URL: <https://aclanthology.org/2024.c3nlp-1> [Accessed 2026-04-01] [doi: [10.18653/v1/2024.c3nlp-1.3](https://doi.org/10.18653/v1/2024.c3nlp-1.3)]
24. Portal site of official statistics of Japan. e-Stat. 2024. URL: <https://www.e-stat.go.jp/en> [Accessed 2024-03-29]
25. Census bureau data. United States Census Bureau. 2024. URL: <https://data.census.gov> [Accessed 2024-03-29]
26. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: cross-sectional survey on COVID-19 in the United States. *J Med Internet Res*. Jun 16, 2020;22(6):e19782. [doi: [10.2196/19782](https://doi.org/10.2196/19782)] [Medline: [32501801](https://pubmed.ncbi.nlm.nih.gov/32501801/)]
27. United States cities database. Simplemaps. 2022. URL: <https://simplemaps.com/data/us-cities> [Accessed 2024-01-02]
28. MIT. List of cities in Japan. GitHub. 2019. URL: <https://github.com/nobuf/list-of-cities-in-japan> [Accessed 2024-01-02]
29. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. Apr 1960;20(1):37-46. [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]
30. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-282. [doi: [10.11613/BM.2012.031](https://doi.org/10.11613/BM.2012.031)]
31. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. *Internet Interv*. May 2015;2(2):183-188. [doi: [10.1016/j.invent.2015.03.005](https://doi.org/10.1016/j.invent.2015.03.005)]
32. Mozetič I, Grčar M, Smailović J. Multilingual Twitter sentiment classification: the role of human annotators. *PLoS ONE*. 2016;11(5):e0155036. [doi: [10.1371/journal.pone.0155036](https://doi.org/10.1371/journal.pone.0155036)] [Medline: [27149621](https://pubmed.ncbi.nlm.nih.gov/27149621/)]
33. Bikku T, Jarugula J, Kongala L, Tummala ND, Vardhani Donthiboina N. Exploring the effectiveness of BERT for sentiment analysis on large-scale social media data. Presented at: 2023 3rd International Conference on Intelligent Technologies (CONIT); Jun 23-25, 2023:1-4; Hubli, India. [doi: [10.1109/CONIT59222.2023.10205600](https://doi.org/10.1109/CONIT59222.2023.10205600)]
34. Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized BERT pretraining approach. corr. Preprint posted online on Jul 26, 2019. URL: <http://arxiv.org/abs/1907.11692>
35. Tohoku NLP. Pretrained Japanese BERT models. Hugging Face. 2023. URL: <https://huggingface.co/tohoku-nlp/bert-base-japanese-v3> [Accessed 2024-05-22]
36. Duarte JM, Berton L. A review of semi-supervised learning for text classification. *Artif Intell Rev*. Sep 2023;56(9):9401-9469. [doi: [10.1007/s10462-023-10393-8](https://doi.org/10.1007/s10462-023-10393-8)]
37. Klein AZ, Kunatharaju S, O'Connor K, Gonzalez-Hernandez G. Automatically identifying self-reports of COVID-19 diagnosis on Twitter: an annotated data set, deep neural network classifiers, and a large-scale cohort. *J Med Internet Res*. Jul 3, 2023;25:e46484. [doi: [10.2196/46484](https://doi.org/10.2196/46484)] [Medline: [37399062](https://pubmed.ncbi.nlm.nih.gov/37399062/)]
38. Klein AZ, Gutiérrez Gómez JA, Levine LD, Gonzalez-Hernandez G. Using longitudinal Twitter data for digital epidemiology of childhood health outcomes: an annotated data set and deep neural network classifiers. *J Med Internet Res*. Mar 25, 2024;26:e50652. [doi: [10.2196/50652](https://doi.org/10.2196/50652)] [Medline: [38526542](https://pubmed.ncbi.nlm.nih.gov/38526542/)]
39. Suzuki R, Iizuka Y, Sugawara H, Lefor AK. Wearing masks is easy but taking them off is difficult - A situation in Japan during COVID-19 pandemic and after. *Dialogues Health*. Jun 2024;4:100172. [doi: [10.1016/j.dialog.2024.100172](https://doi.org/10.1016/j.dialog.2024.100172)] [Medline: [38516216](https://pubmed.ncbi.nlm.nih.gov/38516216/)]
40. Shepardson D, Singh RK, Mason J. US will no longer enforce mask mandate on airplanes, trains after court ruling. Reuters. 2022. URL: <https://www.reuters.com/legal/government/us-judge-rules-mask-mandate-transport-unlawful-overturning-biden-effort-2022-04-18/> [Accessed 2024-05-29]
41. Karako K, Song P, Chen Y, Tang W, Kokudo N. Overview of the characteristics of and responses to the three waves of COVID-19 in Japan during 2020-2021. *BST*. 2021;15(1):1-8. [doi: [10.5582/bst.2021.01019](https://doi.org/10.5582/bst.2021.01019)]

42. Kamba M, She WJ, Ferawati K, Wakamiya S, Aramaki E. Exploring the impact of the COVID-19 pandemic on Twitter in Japan: qualitative analysis of disrupted plans and consequences. *JMIR Infodemiology*. Apr 1, 2024;4:e49699. [doi: [10.2196/49699](https://doi.org/10.2196/49699)] [Medline: [38557446](https://pubmed.ncbi.nlm.nih.gov/38557446/)]
43. Ramos PJ, Ferawati K, Liew K, Aramaki E, Wakamiya S. Emotion analysis of writers and readers of Japanese tweets on vaccinations. Presented at: Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis; May 26, 2022:95-103; Dublin, Ireland. [doi: [10.18653/v1/2022.wassa-1.10](https://doi.org/10.18653/v1/2022.wassa-1.10)]
44. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study. *JMIR Public Health Surveill*. Sep 25, 2018;4(3):e65. [doi: [10.2196/publichealth.8627](https://doi.org/10.2196/publichealth.8627)] [Medline: [30274968](https://pubmed.ncbi.nlm.nih.gov/30274968/)]
45. Liang G, Zhao J, Lau HYP, Leung CK. Using social media to analyze public concerns and policy responses to COVID-19 in Hong Kong. *ACM Trans Manage Inf Syst*. Dec 31, 2021;12(4):1-20. [doi: [10.1145/3460124](https://doi.org/10.1145/3460124)]
46. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. *J Med Internet Res*. 2020;22(5):e18796. [doi: [10.2196/18796](https://doi.org/10.2196/18796)]
47. Liu E, Arledge S. Individual characteristics and demographics associated with mask wearing during the COVID-19 pandemic in the United States. *Asian J Soc Health Behav*. 2022;5(1):3-9. [doi: [10.4103/shb.shb\\_148\\_21](https://doi.org/10.4103/shb.shb_148_21)]
48. Kim JK, Crimmins EM. How does age affect personal and social reactions to COVID-19: results from the national understanding America study. *PLoS ONE*. 2020;15(11):e0241950. [doi: [10.1371/journal.pone.0241950](https://doi.org/10.1371/journal.pone.0241950)] [Medline: [33170903](https://pubmed.ncbi.nlm.nih.gov/33170903/)]
49. Social media fact sheet. Pew Research Center. 2024. URL: <https://www.pewresearch.org/internet/fact-sheet/social-media/> [Accessed 2024-05-29]
50. Howard J, Huang A, Li Z, et al. An evidence review of face masks against COVID-19. *Proc Natl Acad Sci USA*. Jan 26, 2021;118(4):e2014564118. [doi: [10.1073/pnas.2014564118](https://doi.org/10.1073/pnas.2014564118)]
51. Esmailzadeh P. Public concerns and burdens associated with face mask-wearing: Lessons learned from the COVID-19 pandemic. *Prog Disaster Sci*. Jan 2022;13:100215. [doi: [10.1016/j.pdisas.2022.100215](https://doi.org/10.1016/j.pdisas.2022.100215)] [Medline: [35036901](https://pubmed.ncbi.nlm.nih.gov/35036901/)]
52. Sugimura M, Chimed-Ochir O, Yumiya Y, et al. The association between wearing a mask and COVID-19. *Int J Environ Res Public Health*. Aug 30, 2021;18(17):9131. [doi: [10.3390/ijerph18179131](https://doi.org/10.3390/ijerph18179131)] [Medline: [34501719](https://pubmed.ncbi.nlm.nih.gov/34501719/)]
53. Daoust JF. Elderly people and responses to COVID-19 in 27 countries. *PLoS ONE*. 2020;15(7):e0235590. [doi: [10.1371/journal.pone.0235590](https://doi.org/10.1371/journal.pone.0235590)] [Medline: [32614889](https://pubmed.ncbi.nlm.nih.gov/32614889/)]
54. Takahashi Y, Kubota H, Shigeto S, Yoshida T, Yamagata Y. Diverse values of urban-to-rural migration: a case study of Hokuto City, Japan. *J Rural Stud*. Oct 2021;87:292-299. [doi: [10.1016/j.jrurstud.2021.09.013](https://doi.org/10.1016/j.jrurstud.2021.09.013)]
55. Zang E, West J, Kim N, Pao C. U.S. regional differences in physical distancing: evaluating racial and socioeconomic divides during the COVID-19 pandemic. *PLoS One*. 2021;16(11):e0259665. [doi: [10.1371/journal.pone.0259665](https://doi.org/10.1371/journal.pone.0259665)] [Medline: [34847174](https://pubmed.ncbi.nlm.nih.gov/34847174/)]
56. Jacobs P, Ohinmaa AP. The enforcement of statewide mask wearing mandates to prevent COVID-19 in the US: an overview. *F1000Res*. 2020;9:1100. [doi: [10.12688/f1000research.25907.1](https://doi.org/10.12688/f1000research.25907.1)] [Medline: [33520192](https://pubmed.ncbi.nlm.nih.gov/33520192/)]
57. Timpka T, Nyce JM. Face mask use during the COVID-19 pandemic—the significance of culture and the symbolic meaning of behavior. *Ann Epidemiol*. Jul 2021;59:1-4. [doi: [10.1016/j.annepidem.2021.03.012](https://doi.org/10.1016/j.annepidem.2021.03.012)] [Medline: [33836290](https://pubmed.ncbi.nlm.nih.gov/33836290/)]
58. Kemmelmeier M, Jami WA. Mask wearing as cultural behavior: an investigation across 45 U.S. states during the COVID-19 pandemic. *Front Psychol*. 2021;12:648692. [doi: [10.3389/fpsyg.2021.648692](https://doi.org/10.3389/fpsyg.2021.648692)] [Medline: [34366970](https://pubmed.ncbi.nlm.nih.gov/34366970/)]
59. Gotanda H, Miyawaki A, Tabuchi T, Tsugawa Y. Association between trust in government and practice of preventive measures during the COVID-19 pandemic in Japan. *J Gen Intern Med*. Nov 2021;36(11):3471-3477. [doi: [10.1007/s11606-021-06959-3](https://doi.org/10.1007/s11606-021-06959-3)] [Medline: [34159544](https://pubmed.ncbi.nlm.nih.gov/34159544/)]
60. Suppasri A, Kitamura M, Tsukuda H, et al. Perceptions of the COVID-19 pandemic in Japan with respect to cultural, information, disaster and social issues. *Prog Disaster Sci*. Apr 2021;10:100158. [doi: [10.1016/j.pdisas.2021.100158](https://doi.org/10.1016/j.pdisas.2021.100158)] [Medline: [34095808](https://pubmed.ncbi.nlm.nih.gov/34095808/)]
61. Sizing up Twitter users. Pew Research Center. 2019. URL: <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/> [Accessed 2024-07-17]
62. FY2023 survey report on usage time of information and communications media and information behavior. Ministry of Internal Affairs and Communications Japan; 2024. URL: [https://www.soumu.go.jp/main\\_sosiki/joho\\_tsusin/eng/pressrelease/2024/pdf/000382186\\_20240621\\_4.pdf](https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2024/pdf/000382186_20240621_4.pdf) [Accessed 2026-04-01]
63. Pokhriyal N, Valentino BA, Vosoughi S. Quantifying participation biases on social media. *EPJ Data Sci*. 2023;12(1):26. [doi: [10.1140/epjds/s13688-023-00405-6](https://doi.org/10.1140/epjds/s13688-023-00405-6)]
64. Fielding T, Ishikawa Y. COVID-19 and migration: a research note on the effects of COVID-19 on internal migration rates and patterns in Japan. *Popul Space Place*. Aug 2021;27(6):e2499. [doi: [10.1002/psp.2499](https://doi.org/10.1002/psp.2499)] [Medline: [34512211](https://pubmed.ncbi.nlm.nih.gov/34512211/)]

65. Nishimura Y, Ochi K, Tokumasu K, et al. Impact of the COVID-19 pandemic on the psychological distress of medical students in Japan: cross-sectional survey study. *J Med Internet Res.* 2021;23(2):e25232. [doi: [10.2196/25232](https://doi.org/10.2196/25232)]
66. Noda T, Nagaura H, Tsutsumi T, et al. A cross-sectional study of the psychological impact of the COVID-19 pandemic on undergraduate and graduate students in Japan. *J Affect Disord Rep.* Dec 2021;6:100282. [doi: [10.1016/j.jadr.2021.100282](https://doi.org/10.1016/j.jadr.2021.100282)]
67. Smail E, Schneider KE, DeLong SM, et al. Health beliefs and preventive behaviors among adults during the early COVID-19 pandemic in the United States: a latent class analysis. *Prev Sci.* Nov 2021;22(8):1013-1022. [doi: [10.1007/s11121-021-01273-0](https://doi.org/10.1007/s11121-021-01273-0)] [Medline: [34275054](https://pubmed.ncbi.nlm.nih.gov/34275054/)]
68. Fujii R, Suzuki K, Niimi J. Public perceptions, individual characteristics, and preventive behaviors for COVID-19 in six countries: a cross-sectional study. *Environ Health Prev Med.* Mar 3, 2021;26(1):29. [doi: [10.1186/s12199-021-00952-2](https://doi.org/10.1186/s12199-021-00952-2)] [Medline: [33657995](https://pubmed.ncbi.nlm.nih.gov/33657995/)]
69. Xue J, Chen J, Hu R, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. *J Med Internet Res.* Nov 25, 2020;22(11):e20550. [doi: [10.2196/20550](https://doi.org/10.2196/20550)] [Medline: [33119535](https://pubmed.ncbi.nlm.nih.gov/33119535/)]
70. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health Surveill.* Nov 11, 2020;6(4):e21978. [doi: [10.2196/21978](https://doi.org/10.2196/21978)] [Medline: [33108310](https://pubmed.ncbi.nlm.nih.gov/33108310/)]
71. Hu T, Wang S, Luo W, et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. *J Med Internet Res.* Sep 10, 2021;23(9):e30854. [doi: [10.2196/30854](https://doi.org/10.2196/30854)] [Medline: [34346888](https://pubmed.ncbi.nlm.nih.gov/34346888/)]
72. Suarez-Lledo V, Alvarez-Galvez J. Assessing the role of social bots during the COVID-19 pandemic: infodemic, disagreement, and criticism. *J Med Internet Res.* Aug 25, 2022;24(8):e36085. [doi: [10.2196/36085](https://doi.org/10.2196/36085)] [Medline: [35839385](https://pubmed.ncbi.nlm.nih.gov/35839385/)]
73. Chang HCH, Ferrara E. Comparative analysis of social bots and humans during the COVID-19 pandemic. *J Comput Soc Sci.* 2022;5(2):1409-1425. [doi: [10.1007/s42001-022-00173-9](https://doi.org/10.1007/s42001-022-00173-9)] [Medline: [35789937](https://pubmed.ncbi.nlm.nih.gov/35789937/)]

## Abbreviations

**BERT:** Bidirectional Encoder Representations from Transformers

**CDC:** Centers for Disease Control and Prevention

**SoE:** state of emergency

*Edited by Andre Kushniruk; peer-reviewed by KittisaK Jermsttiparsert; submitted 29.May.2025; final revised version received 14.Feb.2026; accepted 20.Feb.2026; published 08.Apr.2026*

### *Please cite as:*

*Ferawati K, She WJ, Wakamiya S, Aramaki E*

*Comparison of Mask-Wearing Behavior on Social Media and Its Relationship With Demographic Characteristics During the COVID-19 Pandemic: Social Media Data Analysis Between the United States and Japan*

*JMIR Hum Factors* 2026;13:e78236

URL: <https://humanfactors.jmir.org/2026/1/e78236>

doi: [10.2196/78236](https://doi.org/10.2196/78236)

© Kiki Ferawati, Wan Jou She, Shoko Wakamiya, Eiji Aramaki. Originally published in *JMIR Human Factors* (<https://humanfactors.jmir.org>), 08.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Human Factors*, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.