Amsterdam University Press

COMPUTATIONAL COMMUNICATION RESEARCH 5.1 (2023) 1–19
HTTPS://DOI.ORG/10.5117/CCR2023.1.8.WANG

Cross-Platform Information Flow and Multilingual Text Analysis: A Comparative Study of Weibo and Twitter Through Deep Learning

Zituo Wang
University of Southern California

Jiayi Zhu
University of Melbourne

Yixuan Xu
Wuhan University

Donggyu Kim
University of Southern California

Dmitri Williams
University of Southern California

Abstract
This study delved into cross-platform information flow and multilingual text analysis by examining social media posts on Weibo and Twitter in Chinese and English. We investigated public opinions about a violent restaurant attack in China that received widespread attention and validated three strategies of Bidirectional Encoder Representations from Transformers (BERT) to classify multilingual social media posts regarding their attitudes, targets, and frames. This study found that there was more criticism than support on Twitter than on Weibo when calling for social justice. When targeting the governments, Weibo users focused more on the local level, while Twitter users focused more on the state level. When framing their opinions, Weibo users focused more on gender violence, while Twitter users focused more on gang violence. These variations within social media posts across platforms were fundamentally influenced by the interruption of transnational information flow as a result of Chinese governance and censorship of the internet. Through the "porous censorship," social media users' autonomy and trust in the government played critical roles in the dynamics between online criticism and authoritarian responsiveness.

Keywords: Information flow, Multilingual text analysis, Social justice, Public opinion, Deep learning

Introduction

Now we are in a network society (Castells, 1996) in which the information flow constitutes the space of flow that dominates our economic, political, and symbolic life. The information flow facilitated by communication technology, as Castells (1996) argued, absorbs the logic and meaning of places and blurs the relationship between architecture and society by creating "time-sharing social practices." However, although the fluid, dynamic, and mobile information flow demonstrates the power of crossing national borders (Shields, 2014), it is controlled and restricted by governments worldwide, for fear of the decline in the regulatory power of states (Goldsmith & Wu, 2006). Strategies of government censorship, such as internet shutdowns (Mare, 2020), internet filtering (Zittrain & Edelman, 2003), as well as regulatory controls (Deibert, 2009) to repress digital activism and online movement (Earl et al., 2022) and divert the attention of citizens (Roberts, 2018), interrupt the transnational information flow.

As we turn our attention to the information flow between China and the world (Lu et al., 2022), it is crucial to examine its impact on disparate but interconnected discussions around the same news and trending topics. For example, why does information flow across national borders despite government censorship? How have attitudes, targets, and frames of user-generated content changed during the information flow, and to what extent are those changes related to the governance and censorship of the internet? More importantly, when analyzing social media posts in different languages, how can the multilingual analysis methods be validated?

On June 10, 2022, a group of men attacked four women at a restaurant in the Chinese city of Tangshan. The video footage triggered both nationwide and worldwide discussions to call for justice by dividing social media users into different attitudes (e.g., support, criticism) toward different targets (e.g., females, local-level governments, etc.) within different frames (e.g., gender violence, gang violence, etc.). This study examines the multilingual discussions on Twitter and the Chinese social media platform Weibo about this news by comparing English and Chinese texts combining human annotation and deep learning.

After human-annotating 5,000 English tweets and 5,000 Chinese Weibo posts, we conducted experiments using three strategies of Bidirectional Encoder Representations from Transformers (BERT) and compared their performance. By validating the three strategies of BERT, we finally classified all multilingual texts (N = 392,448) into different categories to answer our research questions. This study found that there was more criticism than

support on Twitter than on Weibo when calling for social justice. When targeting the governments, Weibo users focused more on the local level, while Twitter users focused more on the state level. When framing their opinions, Weibo users focused more on gender violence, while Twitter users focused more on gang violence. These variations within social media posts across platforms were fundamentally influenced by the interruption of transnational information flow as a result of Chinese governance and censorship of the internet. Through the "porous censorship" (Roberts, 2018), social media users' autonomy and trust in the government played critical roles in the dynamics between online criticism and authoritarian responsiveness.

Literature

One World, Two Platforms: Online Criticism Through the "Porous Censorship"

The Chinese government has been successfully controlling the domestic internet, including social media platforms, by exercising state power through legal (Zheng, 2013) and technical (Earl et al., 2022) means to create accessible but repressive online environments, constructing the "networked authoritarianism" termed by MacKinnon (2011). The censorship techniques implemented by governments and platforms, such as search filtering, keyword blocking, and account deletion (Fu et al., 2013; Jiang, 2014), enable the authoritarian government to respond to threats to regime power (Dal et al., 2023). The voice of dissidents, lawyers, intellectuals, and prominent personalities are targeted (Dimitrov, 2017) since conventional theories on authoritarianism point out that the principal objective of censorship is to suppress criticisms (Geddes & Zaller, 1989). As a result, self-censorship of direct and straightforward criticisms widely exists on Chinese social media to protect users themselves (Luqiu, 2017). Furthermore, to deal with the information flow from outside, the Chinese government selectively blocks internet service and social media platforms (e.g., Facebook, Google, Twitter, etc.) by building up the "Great Firewall" (Ensafi et al., 2015). Thus, Chinese social media users face censorship from domestic platforms such as Weibo, especially when criticizing the government, and difficulties accessing international platforms such as Twitter simultaneously.

Additionally, Weibo users are mostly not exposed to major foreign news outlets that report criticizing topics against the Chinese government, such as human rights and political activism (Lu et al., 2016), since China blocks local access to nearly a quarter of them, including BBC, Bloomberg, The

Economist, The Guardian, The New York Times, Reuters, The Wall Street Journal, and The Washington Post (Davis, 2019). Regarding the sentiment comparison of social media posts on Weibo and Twitter, Gao et al. (2012) have found that Weibo users are more likely to post positive messages than Twitter users, and the probability of posting positive messages on Weibo is 11.8% higher than on Twitter. Thus:

RQ1: What were the characteristics of the discussions about the Tangshan incident on Twitter and Weibo regarding language and frequency?

H1: There was more criticism on Twitter than on Weibo about the Tangshan incident.

Criticizing the Chinese Government Wisely: Targeting the Local Level

Although the most restrictive media and online ecosystems in the world are created in China (Chen & Yang, 2019), some criticisms of the government are allowed since government censorship mainly targets those expressions that call for collective demonstrations (King et al., 2013). It is worth mentioning the dynamics between online criticism and authoritarian responsiveness, in which a relatively high level of trust in the Chinese government (Chen et al., 1997; Chen & Shi, 2001) plays a pivotal role. Social media users trust the central government more than the local government (Li, 2016) to help realize social justice by investigating incidents and punishing local officials. Their "creative use" of political criticism (Wu & Fitzgerald, 2021) indirectly targets the government to avoid getting censored. In response, the government adopts a multifaceted approach to address various challenges it faces, such as collective actions or legitimacy erosions, and cosmetic needs or mundane complaints. To manage political challenges, it employs censorship measures to curtail dissent (Shao, 2018). Otherwise, it tolerates criticisms of the government's performance within certain limits and projects an image of responsiveness to appease the public (Wang & Han, 2023). Consequently, within the dynamics, a higher level of responsiveness from the state toward societal demands leads to a diminished intensity of autonomous social movements (Castells, 2015), resulting in criticisms against the government on Weibo confined in scale and intensity.

On the other hand, as the Chinese government lacks direct control over overseas social media platforms such as Twitter, users who wish to access and share content not permitted on domestic social media have crossed the national (and legal) border using tools such as Virtual Private Networks (VPN). On Twitter, they share technical knowledge to bypass the firewall,

express political opinions, mobilize activism actions, and disseminate alternative news items (Wu & Mai, 2019; Xu & Feng, 2015), which may target the central government without censorship. However, the differences when social media users call for justice for the same incident regarding their specific targets lack research on a micro level. To address this gap, our paper examines the following:

H2: When targeting the governments, Weibo users focused more on the local level, while Twitter users focused more on the state level.

Gender Violence or Gang Violence?

We define the frame of posts as "gender violence" when users describe the incident as a group of males beating a group of females; we define the frame of posts as "gang violence" when users portray the incident as a group of gangsters or bad people beating innocent or good people. The online concentration on gender violence has been fundamentally linked to digital activism, especially when fostering a rising politics of gender justice (see: Dey, 2020; Puente et al., 2021; Williams, 2016). Powered by self-communication technology (Castells, 2007), feminists use social media to express a specific presence or voice that is more difficult to sustain using traditional modes (Wang & Driscoll, 2019). In China, feminists mobilize public support through social media by aliening mainstream journalists and deploying hashtag campaigns (see: Han, 2018; Li & Li, 2017; Wang & Driscoll, 2019). However, due to concerns about the potential consequences of collective actions driven by feminist activism, the Chinese government attempts to contain online feminist activism by removing feminist content and shutting down activists' social media accounts (Fincher, 2016; King et al., 2013). After the incident, Weibo implemented a zero-tolerance policy toward "harmful speech," including posts that "attacked state policy and the political system" or that "incited gender conflict," by removing more than fourteen thousand posts, suspending eight thousand users and permanently banning one thousand users (Zhang, 2022). Additionally, feminist activism in China is perceived as morally deviant, foreign-rooted, and intertwined with nationalism and a modernization-tradition contradiction (Huang, 2016), impacted by the culturally ingrained hegemonic masculinity, structural gender inequality, and the state's pursuit of capitalist economic development (Luo, 2017).

On the contrary, the frame of gang violence was favored by the government and promoted by the state-run media, since the campaign to "sweep away black societies and eradicate evil forces" started in 2018 targeted gang violence and crimes of "black and evil forces" by mobilizing the centralized

law enforcement that underpined China's campaign-style justice (Yin & Mou, 2023). Campaign-style law enforcement strengthened China's authoritarian regime by resolving the legitimacy crisis caused by the economic slowdown, infiltration of gangs into grassroots political structures, and problems of police corruption and shirking (Wang, 2020). As a result of authoritarian responsiveness, the ringleader of the incident was sentenced to 24 years, and eight other people were jailed as a whole "evil force," with authorities framing the incident as a gang-related crime (Reed, 2022). Thus, discussions under the frame of gang violence were less likely to be censored since they were aligned with the government's discourse of "sweeping away black societies and eradicating evil forces" and campaign-style law enforcement targeting gang crimes. Therefore, we hypothesize:

H3: When framing their opinions, Weibo users focused more on gang violence, while Twitter users focused more on gender violence.

Multilingual Text Analysis: Crossing Language Borders

Past studies have shown three main gaps in computational text analysis methods (CTAM) (Baden et al., 2022), especially when dealing with cases that include different languages. To complete multilingual text analysis, keywords-based dictionaries (Dobbrick et al., 2022; Lind et al., 2019) and topic models (Lind et al., 2022; Maier et al., 2022) approaches were used to classify different themes and narratives among texts across language barriers. However, those measures created the gap between CTAMs' tendency to focus on precisely one kind of information and researchers' need for the measurement of multiple and textual contents (Baden et al., 2022), because the main logic of them was extracting specific contents using simple keywords or formatting rules to determine the theme or narrative instead of considering the text as a whole in certain contexts.

Recently, as a machine learning technique, deep learning has facilitated the construction of computational models that consist of multiple layers of processing to learn representations of data such as text with multiple levels of abstraction (LeCun et al., 2015). Bidirectional Encoder Representations from Transformers (BERT), which are developed for pre-training deep bidirectional representations by employing joint conditioning across all layers (Devlin et al., 2018), have been utilized in communication studies as one of the CTAMs. By performing deep learning to achieve a relatively comprehensive understanding of texts as data, BERT can be trained to classify texts in accordance with the distinct demands of specific tasks. For example, Lu et al. (2021) used the Chinese BERT to analyze the public sentiment on Weibo

during the COVID-19 pandemic. Since the use of BERT for multilingual text analysis is still limited, analyzing texts in different languages by including BERT as a new method is worth further testing.

Before the actual step of multilingual text analysis, past research used machine translation to overcome language gaps (Lind et al., 2021; Reber, 2019), which has been proven to be useful for further comparative text analysis (De Vries et al., 2018). But Chan et al. (2020) argued that machine translation should be avoided because it lacks a static version for reproducibility and can be ineffective in contextualizing the understanding of words. Thus, this study uses and validates BERT in different strategies (with and without machine translation) to complete multilingual text analysis. The following research questions are raised:

RQ2: What are the advantages and disadvantages of using BERT for multilingual text analysis?

RQ3: How to compare and validate different strategies using BERT for multilingual text analysis?

Method

Data Collection

This study uses multilingual texts on Weibo and Twitter as data. In the second quarter of 2022, Weibo recorded 252 million daily active users[1] while Twitter recorded 238 million daily active users[2]. Original social media posts related to the restaurant attack from June 10, 2022, to July 31, 2022, were collected from the two platforms. We used Python 3 to collect relevant Weibo posts by searching 15 Weibo topics, such as #Tangshan, #TangshanAttack, #9SuspectsWereAllArrested, under the platform's searching tags of "original" and "trending." We used Twitter's API to collect relevant tweets by searching 26 keywords, such as "Tangshan, attack," "Tangshan, assault," "Tangshan, violence." The post's posting time, the number of likes, replies, reposts, and the user's reported gender (only on Weibo), the number of followings, followers were collected. We collected 595,480 original Weibo posts, which were in Chinese, and got 358,677 after removing the duplicate content. We collected 59,653 original tweets and got 36,049 after removing the duplicate content, from which we selected the 33,771 English tweets tagged by Twitter's language classification system.

[1]Weibo quarterly results: http://ir.weibo.com/financial-information/quarterly-results

[2]Twitter quarterly results: https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/

Text Categories and Human Annotation

To test the hypotheses, we divide the texts into categories from three aspects. The attitudes are divided into two categories: 1. Support; 2. Criticism. The targets of support and criticism are divided into five categories: 1. Females; 2. Males; 3. State-level governments; 4. Local-level governments; 5. Media or social platforms. The frames are divided into five categories: 1. News or information; 2. Gender violence; 3. Gang violence; 4. Cyber violence; 5. Questioning information released or censored by governments or platforms.

We randomly selected 6,000 posts (12,000 in total) from each of the two datasets for manual annotation, of which 1,000 were evenly divided into four groups for reliability testing, and 5,000 were used for final manual annotation. Three native speakers of Chinese with proficiency in English (IELTS 7 or higher) received a week of training to achieve a higher percent agreement. They manually annotated 250 Chinese Weibo posts and 250 English tweets according to the same criteria to determine different attitudes, targets, and frames. After the first round of annotation, the consistency rate of the result was 71%; Three labeling personnel were then trained and tested for consistency in the following rounds. The results of the last three rounds were 79%, 88%, and 96%, respectively. We believe that the coding has reached acceptable reliability and the three coders annotated 5,000 Weibo posts and 5,000 tweets. The final annotations were determined by majority agreements.

Attitudes. By referring to the classification definition by Lu et al. (2021), we classify a post as support if it includes a positive evaluation or projects positive emotions (attitude 1); and classify a post as criticism if it includes a negative evaluation or projects negative emotions (attitude 2). If both attitudes were generated within a single post, we select its main attitude.

Targets. We divide targets into people and institutions including females (target 1), males (target 2), state-level governments (target 3), local-level governments (target 4), and media or social platforms (target 5). State-level governments refer to the central government of China, the Chinese Communist Party, the socialist system, and the state-level leaders, while local-level governments include several tiers from the district level (Lubei District), city level (Tangshan City), to the province level (Hebei Province), such as the local police, or the municipal government. If multiple targets were mentioned within a single post, we select its main target.

Frames. As for frames, our focus centers on how users interpret the incident and express their opinions. Thus, if the post is completely reporting or covering the incident (frame 1) or simply spreading cyber violence (frame

4), we regard them as non-opinion framed posts. For those opinion-framed posts, there are not only different understandings of the incident, such as it was an attack of males attacking females (frame 2) or it was an attack of gangsters beating ordinary people (frame 3), but also questioning information released by the government and state-run media or doubting that truth was censored by governments and platforms (frame 5). If multiple frames were used within a single post, we select its main frame.

Multilingual Text Analysis

Based on human-annotated texts, Bidirectional Encoder Representations from Transformers (BERT) models will be trained and validated in the three following strategies. To run the experiments and compare the strategies, we developed three BERT models including a Chinese BERT model, an English BERT model, and a multilingual BERT model, which can be found at: https://github.com/antalhs/Text-prediction-based-on-BERT-model-Chinese-English-. Among the three strategies, we choose the one with the best performance after conducting experiments running three strategies respectively for each classification task (classifying attitudes, targets, and frames). As the F1 score represents the model's combination of precision and recall, which can be used to reflect how exactly machine-predicted results match human-annotated results, we use the F1 score of the models to evaluate the performance of the strategies. Then, we employ the strategy with the best performance to classify the rest of the dataset.

Train a monolingual model and translate the multiple-language dataset into one language. A: Translate English tweets into Chinese through Google Translate API. Use the 5,000 hand-annotated Weibo posts and 5,000 hand-annotated and translated tweets to train the Chinese BERT with the Whole Word Masking model (Chinese BERT-wwm-ext). This strategy will be validated by the F1 score of the model. **B:** Translate Chinese Weibo posts into English through Google Translate API. Use the 5,000 hand-annotated tweets and 5,000 hand-annotated and translated Weibo posts to train the English BERT with the BERT base model. This strategy will be validated by the F1 score of the model;

Train two models respectively without machine translation. Use the 5,000 hand-annotated Weibo posts to train the Chinese BERT with the Whole Word Masking model (Chinese BERT-wwm-ext), and use the 5,000 hand-annotated tweets to train the English BERT with the BERT base model. This strategy will be validated by the weighted arithmetic mean of the F1 score of both models;

Train a multilingual model without machine translation. Use Multilingual BERT (M-BERT) with the BERT multilingual base model to train all human-annotated data including 5,000 Weibo posts and 5,000 tweets. M-BERT is a pre-trained model consisting of 12 layers of transformers, being trained from monolingual Wikipedia in 104 languages (Devlin et al., 2018), which enables this strategy to avoid machine translation within a single model. This strategy will be validated by the F1 score of the model.

Pearson's Chi-squared Test

After classifying all multilingual social media posts regarding the three aspects of attitudes, targets, and frames, we use Pearson's chi-squared difference test to examine the differences between how social media users call for justice on Weibo and Twitter because it allows us to assess the relationship between categorical variables, such as support/criticism, local-level/state-level, and gender violence/gang violence, in a large dataset of multilingual social media posts.

Results

This study totally collected 655,133 social media posts from Weibo and Twitter, finding that the primary language of Weibo posts was Chinese, and the main language of tweets was English. Therefore, this study focused on multilingual text analysis between Chinese and English. After removing duplicated content and social media posts in other languages, we got 358,677 Weibo posts and 33,771 English tweets. We found that the discussion on Weibo started with a high peak in the first five days, while the discussion on Twitter was relatively at a low popularity level and lacked dramatic changes over time (Figure 1).

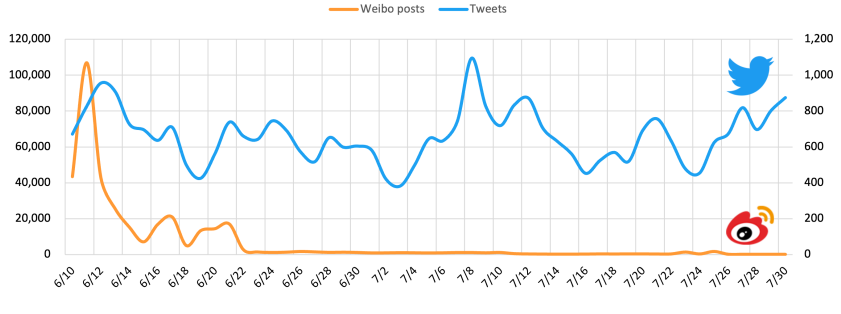

Figure 1: Number of Weibo posts (N = 358,677) and tweets (N = 33,771) from June 10, 2022 to July 31, 2022.

Then, we trained BERT models in three strategies based on the human-annotated texts. Within strategy 1, we used Google Translate API to translate English texts into Chinese texts and translate Chinese texts into English texts in two independent sub-strategies, 1A and 1B. We trained two models of sub-strategies respectively for our three classification tasks, and recorded the F1 scores to measure their performance. For strategy 2, we avoided using machine translation and trained two models for the two languages. After recording the F1 scores of the two models, we used the weighted arithmetic mean of the F1 score of both models, since the amounts of Weibo posts and tweets were not equal. Therefore, we used the final weighted F1 score to evaluate strategy 2. For strategy 3, we simply trained the model without translation and imported all texts into the single model. We consequently recorded the F1 score of the model to evaluate the performance of strategy 3. All F1 scores are demonstrated in Table 1.

Strategies	Attitudes	Targets	Frames	Overall performance
Strategy 1A	0.92	0.74	0.80	0.82
Strategy 1B	0.89	0.78	0.76	0.81
Strategy 2	0.92	0.90	0.83	0.88
Strategy 3	0.90	0.73	0.77	0.80

Table 1: Comparison of performances of BERT strategies

By comparing the F1 scores, strategy 2 achieved the highest performance in three aspects by using two separate models for analyzing texts in two languages. Therefore, we selected strategy 2 to forward our task. Using two models, we predicted the rest of the texts in our dataset. The results are shown in Table 2.

Based on the predicted dataset (N = 392,448), we ran the chi-square difference test finding that the difference between support and criticism on Weibo and Twitter was significant (Pearson Chi-Square Value = 17,630.657, p < .001), and there was more criticism than support on Twitter than on Weibo. Thus, H1 is supported.

We then selected all social media posts targeting the governments (N = 154,003) from Weibo and Twitter to run the chi-square difference test finding that the difference between local-level and state-level on Weibo and Twitter was significant (Pearson Chi-Square Value = 90,512.330, p < .001), and Weibo users focused more on the local-level governments, while Twitter users focused more on the state-level governments. Thus, H2 is supported.

Among those opinion-framed social media posts including gender violence or gang violence (N = 213,568), another chi-square difference test discovered that the difference between frames of gender violence and gang violence on Weibo and Twitter was significant (Pearson Chi-Square Value = 4,562.324, $p < .001$). However, Weibo users focused more on gender violence, while Twitter users focused more on gang violence, opposite to H3.

	Weibo		Twitter	
Attitudes, targets, and frames	Number	Proportion	Number	Proportion
Support	148,398	41%	1,943	6%
Criticism	210,279	59%	31,828	94%
Females	157,436	44%	835	2%
Males	53,636	15%	3,004	9%
State-level governments	15,297	4%	29,625	88%
Local-level governments	108,987	30%	94	0%
Media or social platforms	23,321	7%	213	1%
News or information	72,494	20%	22,097	65%
Gender violence	119,123	33%	1,544	5%
Gang violence	83,375	23%	9,526	28%
Spreading cyber violence	48,375	13%	386	1%
Questioning censorship	35,310	10%	218	1%

Table 2: Comparison of attitudes, targets, and frames of discussions on Weibo and Twitter

Discussion

By analyzing the Chinese and English social media posts, this study explored the differences in what social media users were talking about when calling for justice for a violent incident in China. The differences among their attitudes, targets, and frames demonstrated the fundamental impact on social media users created by the interruption of transnational information flow implemented by the Chinese governance and censorship of the internet.

Weibo, under the Chinese government censorship, and Twitter, uncensored and banned by the Chinese government showed a significant difference in the amount of positive and negative attitudes. On Weibo, substantial posts supported various targets, whereas on Twitter, most posts were criticisms. However, criticisms still existed on Weibo on a large scale, suggesting the operation of "porous censorship" (Roberts, 2018) allowed criticism un-

less it called for collective demonstrations (King et al., 2013). Additionally, Weibo users were not only aware of the censorship (Wang & Mark, 2015), but also raised questions about it using the fifth frame.

By comparing the criticisms across platforms, we revealed different criticizing strategies between Weibo users and Twitter users, who significantly focused more on the local-level governments and the state-level governments, respectively. This discovery was aligned with the high level of trust in the Chinese government (Chen et al., 1997; Chen & Shi, 2001), especially the trust in the central government (Li, 2016), playing a critical role in the dynamics between online criticism and authoritarian responsiveness. On Weibo, social media users demonstrated their autonomy to perceive censorship and avoid censorship by targeting local-level governments instead of state-level governments. Mediated by the trust in the government, it is not surprising that criticism and responsiveness were not in complete opposition but sought each other's support in either achieving social justice or maintaining the role of a "savior" defending people from injustice. On the "uncensored" Twitter, there lacked such an interplay between online criticism and authoritarian responsiveness, resulting in not only criticism becoming the dominating attitude, but also directly targeting state-level governments and institutions.

Our hypothesis was not supported when looking at the frames of user expressions to call for social justice. A much higher proportion of social media posts was framed in gender violence on Weibo, which marked a quite unusual conflict with the state-level discourse of "sweeping away black societies and eradicating evil forces." Nevertheless, as posited by Yang (2014), digital activism would not be rooted out by the adaptability of the Chinese internet control but would develop and adapt to the changing forms of control. The cultural space provided by Chinese social media remained an important means of collective resistance by grassroots prosumers (Mao, 2020). From the censorship side, the nuance of gender violence may not be fully censored by the platform's machine-learning methods since it was not likely to train its censoring model per incident, but we did it in our study so that most posts including (even covert and devious) discussions of gender violence were coded precisely and used to train our own models. From the user side, this finding can also be explained by a relatively high proportion of female users' posts (N = 253,962, 71% of the Weibo dataset) discussing relevant topics on Weibo and a previous discovery of a significant positive relationship between reported gender as female and gender violence frame selection (Wang et al., Working paper). Although we did not have access

to the Twitter user's gender of each tweet, a report[3] showed that 71% of Twitter users were males as of 2022, which was aligned with another previous discovery of a significant positive relationship between reported gender as male and gang violence frame selection (Wang et al., Working paper). On the other hand, Twitter users' criticisms of the state-level policy of "sweeping away black societies and eradicating evil forces" were classified into the gang violence frame, contributing to another possible reason to explain the dominance of this frame over the gender violence frame on Twitter.

This study explored how to use the computational method of deep learning to study large multilingual datasets. In contrast to the existing methods, Bidirectional Encoder Representations from Transformers (BERT) as a deep learning technique, consider the entirety of the text and thereby acquire a deeper understanding of the meaning within its contexts. This technique overcomes the limitations of keywords-based dictionaries and topic models by using thorough deep learning to classify texts for multiple tasks. By validating three strategies of BERT including machine translation in a single model, non-machine translation with separate models, and non-machine translation in a single model, we found that the second strategy achieved the highest performance. As Chan et al. (2020) argued, machine translation should be avoided due to its lack of a static version for reproducibility and its potential ineffectiveness in contextualizing the comprehension of words. In our case, machine translation also demonstrated its shortcoming in performance. As for the M-BERT model, it can be trained without any cross-linguistic target and consistent data (Wang et al., 2019), showcasing surprising cross-linguistic capabilities in previous research. However, the performance was not effective as expected in our case, meaning that M-BERT does not necessarily perform better than BERT models trained based on a specific language (Panda & Levitan, 2021). The reason might be that the certain model included limited pre-trained experience in each language, resulting in relatively low F1 scores.

This study is not without limitations. The data collection was performed during or after the censorship process, creating difficulties in examining a complete dataset of all user's posts and how exactly censorship impacted those posts. Regarding the information flow, it is not clear whether the Weibo users involved in this discussion were the same as those on Twitter or not. As mentioned above, certain key data, such as the users' gender and location, were not available. However, the definite similarity among users on the two platforms was that they shared a common ability to comprehend the

[3]Twitter gender demographics: https://www.oberlo.com/blog/twitter-statistics

incident and express their opinions about it, constructing a strong connection to China-related news. Additionally, because of the limited popularity of this incident, there were rarely third languages talking about it, leaving only Chinese and English to perform multilingual text analysis. Within our two-language dataset, the unbalanced amounts between Chinese Weibo posts and English tweets might have fostered advantages for the Chinese training model to achieve higher performance. Moreover, the text classification using supervised machine learning achieved acceptable performance measured by F1 scores, which can be higher for some subcategories by augmenting human-annotated data to implement further training and tuning the model parameters to optimize specific tasks.

The challenges and potential of applying BERT to analyze multilingual datasets are highlighted, particularly when investigating social media posts in different languages. While BERT has been proven effective in certain situations, it is important to recognize that there is no guarantee that it will behave similarly across other languages, especially morphologically rich languages which encode a lot of information through inflections. The study's limitations, such as the difficulty in examining a complete dataset due to censorship and the focus on only two languages, further underscore the need for caution when generalizing BERT's performance across diverse languages. As a result, researchers must be mindful of these constraints and continue to explore alternative strategies for analyzing multilingual data to ensure accurate and comprehensive insights.

We propose several future research directions to develop the current research. As we have identified three aspects (attitudes, targets, and frames) of social media posts, how did other aspects (e.g., user sentiments) change and evolve over time? In addition to the textual modality, it is imperative to direct focus toward images and videos, since this incident garnered considerable attention and triggered strong indignation initially because of the video footage capturing the violent attack at the restaurant. As for the engagement part, the relationship between social media post metrics (e.g., the number of likes) and their attitudes, targets, and frames are worth exploring to evaluate the effectiveness of users' strategies to call for social justice. The network among the users, including their relationships and interactions, can also be analyzed to determine the out-group and in-group activities and evaluate how such activities factored into the network dynamics of the online community. Furthermore, it remains to be investigated how governments at various levels modified their policies and adjusted their discourse in response to the incident, aiming to mitigate public outrage and rebuild public

trust, and how much state-run media succeeded in helping the government shape public opinions through agenda setting.

This study contributed to the research on cross-platform information flow and multilingual text analysis. By addressing our research questions, we compared different strategies of BERT to analyze the variations within an information flow where social media users' posts called for social justice across language barriers. Based on our dataset (N = 392,448) and 10,000 human-annotated texts, we successfully accomplished the multilingual text analysis task through the validation of three strategies using deep learning. Thus, this study offers valuable insights and practical guidance for future cross-platform and multilingual textual research.

References

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18.

Castells, M. (1996). The information age: Economy, society and culture (3 volumes). *Blackwell, Oxford*, *1997*, 1998.

Castells, M. (2007). Communication, power and counter-power in the network society. *International journal of communication*, *1*(1), 29.

Castells, M. (2015). *Networks of outrage and hope: Social movements in the internet age*. John Wiley & Sons.

Chan, C.-H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., Van Atteveldt, W., & Althaus, S. L. (2020). Reproducible extraction of cross-lingual topics (rectr). *Communication Methods and Measures*, *14*(4), 285–305.

Chen, J., Zhong, Y., & Hillard, J. W. (1997). The level and sources of popular support for china's current political regime. *Communist and Post-Communist Studies*, *30*(1), 45–64.

Chen, X., & Shi, T. (2001). Media effects on political confidence and trust in the people's republic of china in the post-tiananmen period. *East Asia*, *19*(3), 84–118.

Chen, Y., & Yang, D. Y. (2019). The impact of media censorship: 1984 or brave new world? *American Economic Review*, *109*(6), 2294–2332.

Dal, A., Nisbet, E. C., & Kamenchuk, O. (2023). Signaling silence: Affective and cognitive responses to risks of online activism about corruption in an authoritarian context. *new media & society*, *25*(3), 646–664.

Davis, R. (2019). China bars access to nearly a quarter of foreign news websites. https://variety.com/2019/digital/news/china-foreign-news-websites-censorship-1203381682/.

De Vries, E., Schoonvelde, M., & Schumacher, G. (2018). No longer lost in translation: Evidence that google translate works for comparative bag-of-words text applications. *Political Analysis*, *26*(4), 417–430.

Deibert, R. J. (2009). The geopolitics of internet control: Censorship, sovereignty, and cyberspace. *Routledge handbook of Internet politics*, 323–336.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dey, A. (2020). Sites of exception: Gender violence, digital activism, and nirbhaya's zone of anomie in india. *Violence against women*, *26*(11), 1423–1444.

Dimitrov, M. K. (2017). The political logic of media control in china. *Problems of Post-Communism*, *64*(3-4), 121–127.

Dobbrick, T., Jakob, J., Chan, C.-H., & Wessler, H. (2022). Enhancing theory-informed dictionary approaches with "glass-box" machine learning: The case of integrative complexity in social media comments. *Communication Methods and Measures*, *16*(4), 303–320.

Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, *8*(10), eabl8198.

Ensafi, R., Winter, P., Mueen, A., & Crandall, J. R. (2015). Analyzing the great firewall of china over space and time. *Proc. Priv. Enhancing Technol.*, *2015*(1), 61–76.

Fincher, L. H. (2016). China's feminist five. *Dissent*, *63*(4), 84–90.

Fu, K., Chan, C., & Chau., M. (2013). Assessing censorship on microblogs in china: Discriminatory keyword analysis and the real-name registration policy. *IEEE Internet Computing*, *17*(3), 42–50.

Gao, Q., Abel, F., Houben, G.-J., & Yu, Y. (2012). A comparative study of users' microblogging behavior on sina weibo and twitter. *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings 20*, 88–101.

Geddes, B., & Zaller, J. (1989). Sources of popular support for authoritarian regimes. *American Journal of Political Science*, 319–347.

Goldsmith, J., & Wu, T. (2006). Who controls the internet? illusions of a borderless world. *Nova Iorque: Oxford University Press*.

Han, X. (2018). Searching for an online space for feminism? the chinese feminist group gender watch women's voice and its changing approaches to online misogyny. *Feminist Media Studies*, *18*(4), 734–749.

Huang, Y. (2016). War on women: Interlocking conflicts within the vagina monologues in china. *Asian Journal of Communication*, *26*(5), 466–484.

Jiang, M. (2014). The business and politics of search engines: A comparative study of baidu and google's search results of internet events in china. *New media & society*, *16*(2), 212–233.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American political science Review*, *107*(2), 326–343.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Li, J., & Li, X. (2017). Media as a core political resource: The young feminist movements in china. *Chinese Journal of Communication*, *10*(1), 54–71.

Li, L. (2016). Reassessing trust in the central government: Evidence from five national surveys. *The China Quarterly*, *225*, 100–121.

Lind, F., Eberl, J.-M., Eisele, O., Heidenreich, T., Galyga, S., & Boomgaarden, H. G. (2022). Building the bridge: Topic modeling for comparative research. *Communication Methods and Measures*, *16*(2), 96–114.

Lind, F., Eberl, J.-M., Heidenreich, T., & Boomgaarden, H. G. (2019). Computational communication science| when the journey is as important as the goal: A roadmap to multilingual dictionary construction. *International Journal of Communication*, *13*, 21.

Lind, F., Heidenreich, T., Kralj, C., & Boomgaarden, H. G. (2021). Greasing the wheels for comparative communication research: Supervised text classification for multilingual corpora. *Computational Communication Research*, *3*(3).

Lu, Y., Chu, Y., & Shen, F. (2016). Mass media, new technology, and ideology: An analysis of political trends in china. *Global Media and China*, *1*(1-2), 70–101.

Lu, Y., Pan, J., & Xu, Y. (2021). Public sentiment on chinese social media during the emergence of covid-19. *21st Century China Center Research Paper*, *4*.

Lu, Y., Schaefer, J., Park, K., Joo, J., & Pan, J. (2022). How information flows from the world to china. *The International Journal of Press/Politics*, 19401612221117470.

Luo, W. (2017). Television's "leftover" bachelors and hegemonic masculinity in post-socialist china. *Women's Studies in Communication*, *40*(2), 190–211.

Luqiu, L. R. (2017). The cost of humour: Political satire on social media and censorship in china. *Global Media and Communication*, *13*(2), 123–138.

MacKinnon, R. (2011). China's" networked authoritarianism". *J. Democracy*, *22*, 32.

Maier, D., Baden, C., Stoltenberg, D., De Vries-Kedem, M., & Waldherr, A. (2022). Machine translation vs. multilingual dictionaries assessing two strategies for the topic modeling of multilingual text collections. *Communication methods and measures*, *16*(1), 19–38.

Mao, C. (2020). Feminist activism via social media in china. *Asian Journal of Women's Studies*, *26*(2), 245–258.

Mare, A. (2020). Internet shutdowns in africa| state-ordered internet shutdowns and digital authoritarianism in zimbabwe. *International Journal of Communication*, *14*, 20.

Panda, S., & Levitan, S. I. (2021). Detecting multilingual covid-19 misinformation on social media via contextualized embeddings. *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, 125–129.

Puente, S. N., Maceiras, S. D., & Romero, D. F. (2021). Twitter activism and ethical witnessing: Possibilities and challenges of feminist politics against gender-based violence. *Social science computer review*, *39*(2), 295–311.

Reber, U. (2019). Overcoming language barriers: Assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text corpora. *Communication methods and measures*, *13*(2), 102–125.

Reed, B. (2022). China sentences man who attacked women at restaurant to 24 years. https://www.theguardian.com/world/2022/sep/23/china-sentences-man-who-attacked-women-at-restaurant-to-24-years.

Roberts, M. (2018). *Censored: Distraction and diversion inside china's great firewall*. Princeton University Press.

Shao, L. (2018). The dilemma of criticism: Disentangling the determinants of media censorship in china. *Journal of East Asian Studies*, *18*(3), 279–297.

Shields, P. (2014). Borders as information flows and transnational networks. *Global Media and Communication*, *10*(1), 3–33.

Wang, B., & Driscoll, C. (2019). Chinese feminists on social media: Articulating different voices, building strategic alliances. *Continuum*, *33*(1), 1–15.

Wang, D., & Mark, G. (2015). Internet censorship in china: Examining user awareness and attitudes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *22*(6), 1–22.

Wang, P. (2020). Politics of crime control: How campaign-style law enforcement sustains authoritarian rule in china. *The British Journal of Criminology*, *60*(2), 422–443.

Wang, Y., & Han, R. (2023). Cosmetic responsiveness: Why and how local authorities respond to mundane online complaints in china. *Journal of Chinese Political Science*, *28*(2), 187–207.

Wang, Z., Mayhew, S., Roth, D., et al. (2019). Cross-lingual ability of multilingual bert: An empirical study. *arXiv preprint arXiv:1912.07840*.

Williams, S. (2016). # Sayhername: Using digital activism to document violence against black women. *Feminist media studies*, *16*(5), 922–925.

Wu, S., & Mai, B. (2019). Talking about and beyond censorship: Mapping topic clusters in the chinese twitter sphere. *International Journal of Communication*, *13*, 23.

Wu, X., & Fitzgerald, R. (2021). 'hidden in plain sight': Expressing political criticism on chinese social media. *Discourse Studies*, *23*(3), 365–385.

Xu, W. W., & Feng, M. (2015). Networked creativity on the censored web 2.0: Chinese users' twitter-based activities on the issue of internet censorship. *Journal of Contemporary Eastern Asia*, *14*(1).

Yang, G. (2014). Internet activism & the party-state in china. *Daedalus*, *143*(2), 110–123.

Yin, B., & Mou, Y. (2023). Centralized law enforcement in contemporary china: The campaign to "sweep away black societies and eradicate evil forces". *The China Quarterly*, *254*, 366–380.

Zhang, H. (2022). The censorship machine erasing china's feminist movement. https://www.newyorker.com/news/news-desk/the-censorship-machine-erasing-chinas-feminist-movement.

Zheng, H. (2013). Regulating the internet: China's law and practice. *Beijing L. Rev.*, *4*, 37.

Zittrain, J., & Edelman, B. (2003). Internet filtering in china. *IEEE Internet Computing*, *7*(2), 70–77.