List of Abbreviations

AAAI	......	Association for the Advancement of Artificial Intelligence
ACL	......	Association for Computational Linguistics
ACM	......	Association for Computing Machinery
AI	......	Artificial Intelligence
ARR	......	ACL Rolling Review
COVID-19	......	Coronavirus disease 2019
CNN	......	convolutional neutral networks
CSCW	......	Computer-Supported Cooperative Work And Social Computing
GDELT	......	Global Database of Events, Language, and Tone
ICWSM	......	International AAAI Conference on Web and Social Media
IEEE	......	Institute of Electrical and Electronics Engineers
IRB	......	Institutional Review Board
NLP	......	Natural Language Processing
PACM HCI	......	Proceedings of the ACM on Human Computer Interaction
TABARI	......	Textual Analysis by Augmented Replacement Instructions
URL	......	Uniform Resource Locator

	Misinformation: Ginger/Garlic Train on English								Misinformation: Ginger/Garlic Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pre	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
CTBv1	0.880	0.881	0.879	0.881	0.860	0.836	**0.904**	0.919	0.856	0.835	0.883	0.925	**0.878**	**0.884**	**0.876**	**0.879**
CTBv2	**0.886**	**0.889**	**0.888**	**0.886**	0.774	0.765	0.864	0.836	0.841	0.811	0.891	0.911	0.859	0.865	0.858	0.861
BTweet	**0.886**	0.886	0.887	**0.886**	**0.861**	**0.886**	0.840	**0.936**	**0.868**	0.835	**0.917**	**0.928**	0.837	0.861	0.834	0.842
FT	0.824	0.840	0.821	0.828	0.791	0.805	0.780	0.903	0.618	**0.940**	0.588	0.883	0.439	0.778	0.545	0.575

	Misinformation: Hydroxychloroquine Train on English								Misinformation: Hydroxychloroquine Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pre	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
CTBv1	0.797	0.803	0.792	0.861	**0.811**	**0.833**	**0.800**	**0.839**	**0.824**	**0.834**	**0.817**	**0.847**	0.753	**0.835**	0.718	0.856
CTBv2	**0.825**	0.825	**0.825**	0.878	0.793	0.790	0.796	0.814	0.501	0.605	0.550	0.694	0.462	0.723	0.512	0.781
BTweet	0.817	**0.841**	0.800	**0.881**	0.757	0.803	0.740	0.803	0.800	0.812	0.792	0.828	**0.810**	0.823	**0.799**	**0.872**

	Misinformation: Bioweapon Train on English								Misinformation: Bioweapon Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pre	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
CTBv1	**0.850**	0.871	**0.839**	0.827	**0.888**	**0.890**	**0.887**	**0.894**	0.561	0.516	0.637	0.719	0.516	0.469	0.583	0.683
CTBv2	0.840	**0.882**	0.822	**0.861**	0.861	0.861	0.864	0.867	0.726	0.720	0.756	0.811	0.664	0.683	0.688	0.767
BTweet	0.846	0.861	0.836	**0.861**	0.871	0.870	0.876	0.878	0.872	**0.892**	0.862	0.883	**0.791**	**0.858**	**0.773**	**0.825**
XLM-T	0.776	0.786	0.770	0.797	0.877	0.876	0.879	0.883	**0.880**	**0.892**	**0.872**	**0.889**	0.755	0.764	0.750	0.778

Table 6.3: Averages of results for misinformation detection corresponding to best performance models. Note XLM-T corresponds to the mode that processes the translated Chinese text. Please see Section 6.3.2 for more details.

6.4.1 Misinformation detection

From Table 6.3, we can see using automatic translation methods outperforms using original Chinese tweets processed by multi-lingual model methods in cross-lingual cases in general. Additionally, for most misconceptions, CTBv1, CTBv2, and BTweet can achieve the best performance and non-BERT models rarely perform best in terms of the four metrics. Therefore, in practice, it is highly recommended to use CTBv1, CTBv2, and BTweet. In addition, whether transferring from English to Chinese or from Chinese to English, the zero-shot cross-lingual performance are close to the performance of same language performance. With respect to misinformation detection, zero-shot learning can be used in practice in a bidirectional manner between English and Chinese tweets, highlighting potential uses when moderating multi-lingual content.

6.4.2 Stance detection

As can be seen in Table 6.4, the best performance achieved for the stance detection drops compared with the misinformation detection. Still, for most misconceptions, BTweet and CTBv2 achieve the best performance, both of which are recommended to use in practice. Another observation is that zero-shot learning is more effective when

	Stance: Ginger/Garlic Train on English								Stance: Ginger/Garlic Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
CTBv1	0.693	0.712	0.682	0.775	**0.751**	**0.702**	**0.830**	0.889	0.472	0.443	0.517	0.886	0.521	0.554	0.530	0.711
CTBv2	**0.765**	**0.776**	**0.763**	**0.800**	0.712	0.699	0.741	0.875	0.650	**0.764**	0.649	**0.914**	**0.591**	0.687	0.573	**0.744**
BTweet	0.671	0.665	0.681	0.750	0.733	0.692	0.797	0.881	**0.657**	0.623	**0.708**	0.886	0.588	0.608	**0.581**	0.708
CNN	0.556	0.631	0.537	0.700	0.508	0.635	0.469	**0.897**	0.446	0.433	0.471	0.867	0.432	0.611	0.454	0.589
FT	0.627	0.765	0.581	0.711	0.392	0.419	0.383	0.861	0.360	0.521	0.358	0.897	0.338	**0.707**	0.380	0.628
	Stance: Hydroxychloroquine Train on English								Stance: Hydroxychloroquine Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
CTBv1	0.772	0.772	0.777	0.775	0.702	0.710	**0.705**	**0.731**	0.562	0.544	0.605	0.667	0.455	0.441	0.510	0.528
CTBv2	**0.795**	**0.811**	**0.792**	**0.800**	**0.705**	**0.712**	**0.705**	0.728	0.363	0.364	0.414	0.536	0.302	0.273	0.373	0.408
BTweet	0.707	0.719	0.711	0.711	0.674	0.710	0.668	0.711	**0.669**	**0.711**	**0.659**	**0.742**	**0.620**	**0.640**	**0.629**	**0.625**
	Stance: Bioweapon Train on English								Stance: Bioweapon Train on Chinese							
	Train on English & Test on English				Train on English & Test on Chinese				Train on Chinese & Test on Chinese				Train on Chinese & Test on English			
	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc	F1	Pr	Re	Acc
BTweet	**0.751**	**0.780**	0.739	**0.778**	**0.757**	**0.764**	**0.758**	**0.808**	**0.770**	**0.752**	**0.799**	**0.797**	**0.664**	**0.661**	**0.730**	**0.700**
XLM-T	0.744	0.734	**0.765**	0.767	0.639	0.663	0.635	0.708	0.618	0.622	0.641	0.703	0.599	0.585	0.677	0.619
XLM-T-Original	**0.751**	0.746	0.757	0.758	0.439	0.494	0.464	0.617	0.624	0.630	0.644	0.706	0.479	0.504	0.490	0.547

Table 6.4: Averages of results for stance detection corresponding to best performance models. Note XLM-T and XLM-T-Original correspond to the modes that process the translated and original Chinese text, respectively. Please see Section 6.3.2 for more details.

transfer from English to Chinese than vice versa. Although this is a drawback, it is still possible for zero-shot learning to be used in moderation platforms like Twitter since it is likely that moderators are more familiar with English than Chinese.

6.4.3 Error Analysis

I follow the practice in (Glandt et al., 2021) and conduct a qualitative error analysis to help readers better understand the results.

I choose the case in stance detection related to “ginger/garlic,” and use one of the best models CTBv2 to demonstrate how the model performs. For each test tweet, it can be predicted by models trained on the same language as well as on the cross-lingual manner and all the Chinese tweets mentioned here are translated automatically.

Examples are shown in Table 6.5. Typically, CTBv2 performs well when the stance towards the misconception is none as can be seen in tweet No.1 and No.3. However, CTBv2 stumbles when the meaning of a tweet is vague. One such example is tweet No.2. The human annotators label it as a tweet supports the efficacy of garlic in treating COVID-19 probably because the tweet mentions *the effect is very good*, which the human annotators believe it refers to the effect of garlic. However, one can also

No.	Tweet	Label	$Pred_{zh\_zh}$	$Pred_{en\_zh}$
1	@(username) cry, cry, cry! why is there no iced one, i like it the most! minced garlic and egg yolks are also good! it's all because of the pandemic!	None	None	None
2	@(username) in fact, there are gauze materials that can be used to make masks by yourself.china has a population of 1.4 billion, which cannot be produced and consumes resources. it should teach people all over the country to make masks at home on tv. masks can be sandwiched, dry tea leaves or wormwood leaves/dried garlic chips wait, the effect is very good, and it can block virus droplets.	Support	None	None
No.	Tweet	Label	$Pred_{en\_en}$	$Pred_{zh\_en}$
3	ginger loves covid and rapists	None	None	None
4	my mom think ginger tea gone keep me from getting covid lmaaoo	Refute	Support	Support

Table 6.5: Error analysis for ginger/garlic stance examples. Tweets No.1 and 2 are translated from Chinese. I hide usernames to protect their privacy. $Pred_{zh\_zh}$, $Pred_{en\_zh}$, $Pred_{en\_en}$, and $Pred_{zh\_en}$ stand for the predicted label obtained in the "train on Chinese & test on Chinese", "train on English & test on Chinese", "train on English & test on English", and "train on Chinese & test on English" manners, respectively.

argue that this could refer to the effect of masks. Such controversial tweet prevents the model from predicting correctly. Another potential reason leads to an erroneous prediction is online slang and its variant. As seen in tweet No.4, lmaaoo is a variant of lmao (Dictionary, 2018), showing the author of the tweet that they do not believe ginger is a cure for COVID-19. The variant of this slang may be distant even to a pretrained model, making the model predict incorrectly.

6.5 Discussion

In this section, I discuss the implication, the limitations, potential risks, and privacy issues of this research.

6.5.1 Implications

The results of experiments show CTBv2 and BTweet, i.e., COVID-Tweet-BERT v2 and BERTweet, are generally capable of detecting misconceptions expressed in a tweet and detecting the stance of the author toward this misconception when used in both monolingual and multi-lingual manners. By applying these models, content moderators may pinpoint tweets that are likely to spread certain specific misconceptions, making

3. Do not use information learned from the other tweets.

4. Please take the meaning of the hashtags, mentioned accounts into consideration when reading the content.

5. If the tweet contains non-English, please just decide based on the English part

6. After you submit your answer, you are more than welcome to continue work on the remaining HITs in this project! Thank you very much for your help.

C.4 Detailed Instructions & Examples

1. Does this tweet explicitly or implicitly talk about hydroxychloroquine/chloroquine (HCQ for short) as a treatment or potential treatment of COVID-19?

A. Yes

B. No (if no, please select "Not applicable" for the following two questions)

Instruction & Examples are shown in Table C.1 Add tables.

Instructions	Examples
You should answer "Yes" as long as the tweet mentions the information; that is, even if the tweet refutes the statement that hydroxychloroquine/chloroquine can treat COVID-19, you should still answer yes.	"Repeated studies show #Hydroxychloroquine doesnt work for #COVID19 patients"
You should answer No if the tweet does NOT mention hydroxychloroquine/chloroquine can treat COVID-19 or mentions hydroxychloroquine/chloroquine can treat other diseases.	Hydroxychloroquine is effective against non-resistant strains of Malaria. It has long been known to cause cardiac arrests, but generally thats better than malaria!

Table C.1: Instructions & examples for the first question.

Instructions	Examples
You should answer "Support" when the tweet support the use of hydroxychloroquine/chloroquine as an effective (or potentially effective) treatment of COVID-19 for the general public	We can go support to work; if you get the virus doctors should treat you with hydroxychloroquine. #COVID19
You should answer "Refute" when the tweet does NOT support the use of hydroxychloroquine/chloroquine as an effective (or potentially effective) treatment of COVID-19 for the general public	"Dr. Fauci, an immunologist & Trump's chief at NIAID, says hydroxychloroquine IS NOT effective in preventing coronavirus
You should answer "None" when the tweet has no clear attitude, just jokes around, or cites an objective description without commenting	"Dr. Brian Tysons First-Person Account of Treating COVID-19 with Hydroxychloroquine The Economic Standard"

Table C.2: Instructions & examples for the second question.

2. Considering the overall attitude of the author, does this tweet support or refute the use of hydroxychloroquine/chloroquine as an effective (or potentially effective) treatment of COVID-19 for the general public?

A. Support

B. Refute

C. None

D. Not applicable

Instruction & Examples are shown in Table C.2

3. Does this tweet associate the use/non-use of hydroxychloroquine/chloroquine and COVID-19 and some secret plots by powerful actors, such as governments, politicians, companies (e.g., pharmacies), public figures (e.g., Anthony Fauci or Bill Gates), or other organizations (e.g., CDC, FDA), etc.?

A. Yes

B. No

C. Not applicable

Instruction & Examples are shown in Table C.3

Instructions	Examples
You should answer Yes if the tweet associates the use/non-use of hydroxychloroquine/chloroquine and COVID-19 and some secret plots by powerful actors, such as governments, politicians, companies (e.g., pharmacies), public figures (e.g., Anthony Fauci or Bill Gates), or other organizations (e.g., CDC, FDA), etc.?	In case you are wondering why #Hydroxychloroquine isn't universally being used? Big Pharma makes no money.
You should answer No if there is no such association	"Patients with rheumatic disease who were taking #hydroxychloroquine had a lower risk of #COVID19 infection than patients taking other disease-modifying anti-rheumatic drugs"

Table C.3: Instructions & examples for the third question.

References

Lawrence Hurley (2018). U.S. top court upholds Trump travel ban targeting Muslim-majority nations. https://www.reuters.com/article/us-usa-court-immigration/u-s-top-court-backs-trump-on-travel-ban-targeting-muslim-majority-nations-idUSKBN1JM1U9.

Abidin, C. (2020). Meme factory cultures and content pivoting in Singapore and Malaysia during COVID-19. *Harvard Kennedy School Misinformation Review*.

Ajao, O., Hong, J., and Liu, W. (2015). A survey of location inference techniques on twitter. *Journal of Information Science*, 41(6):855–864.

Akyürek, A. F., Guo, L., Elanwar, R., Ishwar, P., Betke, M., and Wijaya, D. T. (2020). Multi-label and multilingual news framing analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Alam, F., Cresci, S., Chakraborty, T., Silvestri, F., Dimitrov, D., Martino, G. D. S., Shaar, S., Firooz, H., and Nakov, P. (2021). A survey on multimodal disinformation detection. *arXiv preprint arXiv:2103.12541*.

Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G. D. S., Abdelali, A., Durrani, N., Darwish, K., et al. (2020). Fighting the covid-19 infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. *arXiv preprint arXiv:2005.00033*.

Ali, S., Razi, A., Kim, S., Alsoubai, A., Gracie, J., De Choudhury, M., Wisniewski, P. J., and Stringhini, G. (2022). Understanding the digital lives of youth: Analyzing media shared within safe versus unsafe private conversations on instagram. In *CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Aliapoulios, M., Papasavva, A., Ballard, C., De Cristofaro, E., Stringhini, G., Zannettou, S., and Blackburn, J. (2021). The gospel according to q: Understanding the qanon conspiracy from the perspective of canonical information. *arXiv preprint arXiv:2101.08750*.

Allcott, H., Gentzkow, M., and Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2).

Alvari, H. and Shakarian, P. (2019). Hawkes Process for Understanding the Influence of Pathogenic Social Media Accounts. In *arXiv:1902.01970*.

Amy Lange (2016). Detroit family caught in iraq travel ban, says mom died waiting to come home. `https://www.fox5dc.com/news/detroit-family-caught-in-iraq-travel-ban-says-mom-died-waiting-to-come-home`.

Babaei, M., Kulshrestha, J., Chakraborty, A., Benevenuto, F., Gummadi, K. P., and Weller, A. (2018). Purple feed: Identifying high consensus news posts on social media. In *AIES*.

Backstrom, L., Sun, E., and Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70.

Banai, I. P., Banai, B., and Mikloušić, I. (2020). Beliefs in covid-19 conspiracy theories predict lower level of compliance with the preventive measures both directly and indirectly by lowering trust in government medical officials.

Barbieri, F., Anke, L. E., and Camacho-Collados, J. (2021). Xlm-t: A multilingual language model toolkit for twitter. *arXiv preprint arXiv:2104.12250*.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., and Blackburn, J. (2020). The Pushshift Reddit Dataset. In *ICWSM*.

Bayar, B. and Stamm, M. C. (2016). A deep learning approach to universal image manipulation detection using a new convolutional layer. In *ACM IH*.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan).

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10).

Boudemagh, E. and Moise, I. (2017). News media coverage of refugees in 2016: A gdelt case study. In *ICWSM*.

Bozarth, L. and Budak, C. (2020). Toward a better performance evaluation framework for fake news classification. In *ICWSM*.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Brennen, J. S., Simon, F., Howard, P. N., and Nielsen, R. K. (2020). Types, sources, and claims of covid-19 misinformation. *Reuters Institute*, 7(3):1.

Brennen, J. S., Simon, F. M., and Nielsen, R. K. (2021). Beyond (mis) representation: visuals in covid-19 misinformation. *The International Journal of Press/Politics*, 26(1):277–299.

Buchner, J. (2020). A python perceptual image hashing module: Imagehash. `https://github.com/JohannesBuchner/imagehash`. Accessed: 2021-04-08.

Budak, C. (2019). What happened? the spread of fake news publisher content during the 2016 us presidential election. In *The WebConf*.

Castillo, C., Mendoza, M., and Poblete, B. (2011). Information credibility on twitter. In *WWW*.

Chadwick, A. (2011). The hybrid media system. In *ECPR*.

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., and Vakali, A. (2017). Mean birds: Detecting aggression and bullying on twitter. In *Proceedings of the 2017 ACM on web science conference*, pages 13–22.

Chen, E., Lerman, K., and Ferrara, E. (2020a). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.

Chen, K., Chen, A., Zhang, J., Meng, J., and Shen, C. (2020b). Conspiracy and debunking narratives about covid-19 origination on chinese social media: How it started and who is to blame. *arXiv preprint arXiv:2011.08409*.

Chen, K., Duan, Z., and Yang, S. (2022a). Twitter as research data: Tools, costs, skill sets, and lessons learned. *Politics and the Life Sciences*, 41(1):114–130.

Chen, N., Chen, X., Zhong, Z., and Pang, J. (2022b). " double vaccinated, 5g boosted!": Learning attitudes towards covid-19 vaccination from social media. *arXiv preprint arXiv:2206.13456*.

Chicago Tribune (2016). Trump revealed highly classified information to Russian diplomats, U.S. officials say. `http://www.chicagotribune.com/news/nationworld/ct-trump-revealed-classified-information-russians-20170515-story.html`.

Common Crawl Repository (2019).

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., and Menczer, F. (2011a). Predicting the political alignment of twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 192–199. IEEE.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011b). Political polarization on twitter. In *ICWSM*.

Cooper, S. (2007). A Concise History of the Fauxtography Blogstorm in the 2006 Lebanon War. *American Communication Journal*, 9.

Cortis, K. and Davis, B. (2021). A dataset of multidimensional and multilingual social opinions for maltas annual government budget. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 971–981.

Daisuke Wakabayashi, Davey Alba, and Marc Tracy (2020). Bill Gates, at Odds With Trump on Virus, Becomes a Right-Wing Target. `https://www.nytimes.com/2020/04/17/technology/bill-gates-virus-conspiracy-theories.html`.

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

Dewan, P., Suri, A., Bharadhwaj, V., Mithal, A., and Kumaraguru, P. (2017). Towards Understanding Crisis Events On Online Social Networks Through Pictures. In *ASONAM*.

Dictionary, U. (2018). `https://www.urbandictionary.com/define.php?term=lmao`.

Du, Y., Masood, M. A., and Joseph, K. (2020). Understanding visual memes: An empirical analysis of text superimposed on memes shared on twitter. In *ICWSM*.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD*.

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22(8).

Ferrara, E. (2020). What types of covid-19 conspiracies are populated by twitter bots? *arXiv preprint arXiv:2004.09531*.

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., and Moran, S. (2018). Falling for fake news: investigating the consumption of news via social media. In *ACM CHI*.

Flores-Saviaga, C. I., Keegan, B. C., and Savage, S. (2018). Mobilizing the trump train: Understanding collective action in a political trolling community. In *ICWSM*.

Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., Jenner, L., Teale, A.-L., Carr, L., Mulhall, S., et al. (2020). Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in england. *Psychological medicine*, pages 1–13.

Fung, Y. R., Huang, K.-H., Nakov, P., and Ji, H. (2022). The battlefront of combating misinformation and coping with media bias. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4790–4791.

Garber, M. (2017). Al Franken, That Photo, and Trusting the Women. `https://www.theatlantic.com/entertainment/archive/2017/11/al-franken-thatc-and-trusting-the-women/545954/`. Accessed: 2021-04-08.

Garimella, K. and Eckles, D. (2020). Images and misinformation in political groups: Evidence from whatsapp in india. *arXiv:2005.09784*.

Garimella, K., Smith, T., Weiss, R., and West, R. (2021). Political polarization in online news consumption. *arXiv:2104.06481*.

GDELT (2015). The GDELT Event Database Data Format Codebook V2.0.

Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2).

Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2015). Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1.

Glandt, K., Khanal, S., Li, Y., Caragea, D., and Caragea, C. (2021). Stance detection in covid-19 tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, volume 1.

Google (2022). `https://cloud.google.com/translate`.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425).

Guo, F., Blundell, C., Wallach, H., and Heller, K. (2015). The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics*.

Guo, L., Mays, K., Lai, S., Jalal, M., Ishwar, P., and Betke, M. (2020). Accurate, fast, but not always cheap: Evaluating crowdcoding as an alternative approach to analyze social media data. *Journalism & Mass Communication Quarterly*, 97(3):811–834.

Guo, L. and Vargo, C. (2020). fake news and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 us presidential election. *Communication Research*, 47(2):178–200.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., and Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 93(2):332–359.

Guo, L. and Zhang, Y. (2020). Information flow within and across online media platforms: An agenda-setting analysis of rumor diffusion on news websites, weibo, and wechat in china. *Journalism Studies*, 21(15):2176–2195.

Han, B., Cook, P., and Baldwin, T. (2014). Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500.

Hardalov, M., Arora, A., Nakov, P., and Augenstein, I. (2021). A survey on stance detection for mis-and disinformation identification. *arXiv preprint arXiv:2103.00242*.

Hardalov, M., Arora, A., Nakov, P., and Augenstein, I. (2022). Few-shot cross-lingual stance detection with sentiment-based pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 36.

Harder, R. A., Sevenans, J., and Van Aelst, P. (2017). Intermedia agenda setting in the social media age: How traditional players dominate the news agenda in election times. *The International Journal of Press/Politics*, 22(3):275–293.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*.

Hine, G. E., Onaolapo, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Samaras, R., Stringhini, G., and Blackburn, J. (2017). Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web. In *ICWSM*.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hossain, T., Logan IV, R. L., Ugarte, A., Matsubara, Y., Young, S., and Singh, S. (2020). Covidlies: Detecting covid-19 misinformation on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.

Hou, Y., van der Putten, P., and Verberne, S. (2022). The covmis-stance dataset: Stance detection on twitter for covid-19 misinformation. *arXiv preprint arXiv:2204.02000*.

Hu, Y., Huang, H., Chen, A., and Mao, X.-L. (2020). Weibo-cov: A large-scale covid-19 social media dataset from weibo. *arXiv preprint arXiv:2005.09174*.

Infowars (2018). Mexico Agrees to Pay for Wall. `https://www.infowars.com/mexico-agrees-to-pay-for-wall/`.

Javed, R. T., Shuja, M. E., Usama, M., Qadir, J., Iqbal, W., Tyson, G., Castro, I., and Garimella, K. (2020). A first look at covid-19 messages on whatsapp in pakistan. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 118–125. IEEE.

Javed, R. T., Usama, M., Iqbal, W., Qadir, J., Tyson, G., Castro, I., and Garimella, K. (2022). A deep dive into covid-19-related messages on whatsapp in pakistan. *Social Network Analysis and Mining*, 12(1):1–16.

Jiang, S., Metzger, M., Flanagin, A., and Wilson, C. (2020). Modeling and measuring expressed (dis) belief in (mis) information. In *ICWSM*.

JOHN HAYWARD (2017). Seven Inconvenient Facts About Trumps Refugee Actions. `https://www.breitbart.com/politics/2017/01/29/trumps-immigration-pause-sober-defenses-vs-hysterical-criticism/`.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., and Mikolov, T. (2016a). Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2016b). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017). Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431. Association for Computational Linguistics.

Jurgens, D. (2013). That's what friends are for: Inferring location in online social media platforms based on social relationships. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Kar, D., Bhardwaj, M., Samanta, S., and Azad, A. P. (2020). No rumours please! a multi-indic-lingual approach for covid fake-tweet detection. In *2021 Grace Hopper Celebration India (GHCI)*, pages 1–5. IEEE.

Kazemi, A., Garimella, K., Gaffney, D., and Hale, S. A. (2021). Claim matching beyond english to scale global fact-checking. *arXiv preprint arXiv:2106.00853*.

Kiela, D., Firooz, H., Mohan, A., Goswami, V., Singh, A., Fitzpatrick, C. A., Bull, P., Lipstein, G., Nelli, T., Zhu, R., et al. (2021). The hateful memes challenge: competition report. In *NeurIPS 2020 Competition and Demonstration Track*, pages 344–360. PMLR.

Kim, M. G., Kim, M., Kim, J. H., and Kim, K. (2022). Fine-tuning bert models to classify misinformation on garlic and covid-19 on twitter. *International Journal of Environmental Research and Public Health*, 19(9):5126.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Kong, Q. (2019). Linking Epidemic Models and Hawkes Point Processes for Modeling Information Diffusion. In *WSDM*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.

Küçük, D. and Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.

Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., and Karahalios, K. (2019). Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal*, 22(1):188–227.

Kumar, S. and Shah, N. (2018). False information on web and social media: A survey. In *arXiv:1804.08559*.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600.

Lafferty, J., McCallum, A., and Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*.

Lai, M., Cignarella, A. T., Farías, D. I. H., Bosco, C., Patti, V., and Rosso, P. (2020). Multilingual stance detection in social media political debates. *Computer Speech & Language*, 63:101075.

Lazer, D., Ruck, D. J., Quintana, A., Shugars, S., Joseph, K., Grinberg, N., Gallagher, R. J., Horgan, L., Gitomer, A., Bajak, A., et al. (2021). The covid states project# 18: Fake news on twitter.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380).

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

Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., and Satyanarayan, A. (2021). Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–18.

Leetaru, K. and Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone, 1979-2012. In *ISA Annual Convention*.

Leng, Y., Zhai, Y., Sun, S., Wu, Y., Selzer, J., Strover, S., Zhang, H., Chen, A., and Ding, Y. (2021). Misinformation during the covid-19 outbreak in china: cultural, social and political entanglements. *IEEE Transactions on Big Data*, 7(1):69–80.

Leskovec, J., Backstrom, L., and Kleinberg, J. M. (2009). Meme-tracking and the Dynamics of the News Cycle. In *KDD*.

Li, Y. and Xie, Y. (2020). Is a picture worth a thousand words? an empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1):1–19.

Lin, H., Ma, J., Chen, L., Yang, Z., Cheng, M., and Chen, G. (2022). Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. *NAACL 2022*.

Linderman, S. W. and Adams, R. P. (2014). Discovering Latent Network Structure in Point Process Data. In *ICML*.

Linderman, S. W. and Adams, R. P. (2015). Scalable Bayesian Inference for Excitatory Point Process Networks. In *arXiv:1507.03228*.

Lindgren, B. (1993). *Statistical Theory*, volume 22.

Ling, C., AbuHilal, I., Blackburn, J., De Cristofaro, E., Zannettou, S., and Stringhini, G. (2021). Dissecting the meme magic: Understanding indicators of virality in image memes. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–24.

Ling, C., Blackburn, J., De Cristofaro, E., and Stringhini, G. (2022). Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In *14th ACM Web Science Conference 2022*, pages 164–173.

Lucas Ou-Yang (2020). Newspaper3k: Article scraping & curation. `https://newspaper.readthedocs.io/en/latest/`. Accessed: 2021-04-08.

Luceri, L., Giordano, S., and Ferrara, E. (2020). Detecting troll behavior via inverse reinforcement learning: A case study of russian trolls in the 2016 us election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 417–427.

Lukasik, M., Srijith, P., Vu, D., Bontcheva, K., Zubiaga, A., and Cohn, T. (2016). Hawkes processes for continuous time sequence classification: an application to rumour stance classification in twitter. In *ACL*.

Ma, J., Gao, W., and Wong, K.-F. (2017). Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 708–717, Vancouver, Canada. Association for Computational Linguistics.

Ma, Q. and Olshevsky, A. (2020). Adversarial crowdsourcing through robust rank-one matrix completion. *Advances in Neural Information Processing Systems*, 33:21841–21852.

Majestic (2019). The Majestic Million List. `https://majestic.com/reports/majestic-million`. Accessed: 2019-01-28.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *ACL*.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*.

Marcelino, G., Semedo, D., Mourão, A., Blasi, S., Mrak, M., and Magalhaes, J. (2019). A benchmark of visual storytelling in social media. In *ICMR*.

McClatchy DC Bureau (2017). House majority leader told his colleagues in 2016: 'I think Putin pays' Trump. `https://www.mcclatchydc.com/news/politics-government/article151133157.html`.

McGee, J., Caverlee, J., and Cheng, Z. (2013). Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 459–468.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

NewsGuard (2019a). Inside NewsGuard's First Year Fighting Misinformation.

NewsGuard (2019b). Rating Process and Criteria.

NewsGuard (2019c). Sample nutrition labels.

NewsGuard (2019d). The Internet Trust Tool. `https://www.newsguardtech.com/`. Accessed: 2021-04-08.

Ng, L. H. X. and Carley, K. M. (2022). Is my stance the same as your stance? a cross validation study of stance detection datasets. *Information Processing & Management*, 59(6):103070.

Ng, L. H. X., Moffitt, J., and Carley, K. M. (2022). Coordinated through aweb of images: Analysis of image-based influence operations from china, iran, russia, and venezuela. *arXiv preprint arXiv:2206.03576*.

Nguyen, D. Q., Vu, T., and Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 9–14, Online. Association for Computational Linguistics.

Nic Fildes, Mark Di Stefano, and Hannah Murphy (2020). How a 5g coronavirus conspiracy spread across europe. `https://www.ft.com/content/1eeedb71-d9dc-4b13-9b45-fcb7898ae9e1`.

Nørregaard, J., Horne, B. D., and Adalı, S. (2019). Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 630–638.

Oliver, J. E. and Wood, T. (2014). Medical conspiracy theories and health behaviors in the united states. *JAMA internal medicine*, 174(5):817–818.

Papasavva, A., Blackburn, J., Stringhini, G., Zannettou, S., and Cristofaro, E. D. (2021). is it a qoincidence?: An exploratory study of qanon on voat. In *Proceedings of the Web Conference 2021*, pages 460–471.

Papasavva, A., Zannettou, S., De Cristofaro, E., Stringhini, G., and Blackburn, J. (2020). Raiders of the Lost Kek: 3.5 Years of Augmented 4chan Posts from the Politically Incorrect Board. In *ICWSM*.

Park, C. Y., Yan, X., Field, A., and Tsvetkov, Y. (2020). Multilingual contextual affective analysis of lgbt people portrayals in wikipedia. *arXiv preprint arXiv:2010.10820*.

Park, C. Y., Yan, X., Field, A., and Tsvetkov, Y. (2021). Multilingual contextual affective analysis of lgbt people portrayals in wikipedia. In *ICWSM*, pages 479–490.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Paudel, P., Blackburn, J., De Cristofaro, E., Zannettou, S., and Stringhini, G. (2022). Lambretta: Learning to rank for twitter soft moderation. *arXiv preprint arXiv:2212.05926*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *EMNLP*.

Pennycook, G. and Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *PNAS*, 116(7).

Pennycook, G. and Rand, D. G. (2021). The psychology of fake news. *Trends in cognitive sciences*.

Pfeffer, J., Mayer, K., and Morstatter, F. (2018). Tampering with twitters sample api. *EPJ Data Science*, 7(1):50.

Phillips, S. C., Ng, L. H. X., and Carley, K. M. (2022). Hoaxes and hidden agendas: A twitter conspiracy theory dataset: Data paper. In *Companion Proceedings of the Web Conference 2022*, pages 876–880.

Poddar, S., Mondal, M., Misra, J., Ganguly, N., and Ghosh, S. (2022). Winds of change: Impact of covid-19 on vaccine-related opinions of twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 782–793.

Poynter (2020). `https://www.poynter.org/?ifcn_misinformation=studies-show-the-coronavirus-was-engineered-to-be-a-bioweapon`.

Qu, J., Li, L. H., Zhao, J., Dev, S., and Chang, K.-W. (2022). Disinfomeme: A multimodal dataset for detecting meme intentionally spreading out disinformation. *arXiv preprint arXiv:2205.12617*.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., and Menczer, F. (2011). Truthy: mapping the spread of astroturf in microblog streams. In *WWW Companion*.

Ratkiewicz, J., Conover, M., Meiss, M. R., Goncalves, B., Patil, S., Flammini, A., and Menczer, F. (2010). Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams. In *arXiv:1011.3768*.

Reimann, N. (2020). Some americans are tragically still drinking bleach as a coronavirus cure. `https://www.forbes.com/sites/nicholasreimann/2020/08/24/some-americans-are-tragically-still-drinking-bleach-as-a-coronavirus-cure/?sh=421223aa6748`.

Reis, J. C., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., and Benevenuto, F. (2020). A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 903–908.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019). (mis) information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*, pages 818–828.

Resnick, P., Ovadya, A., and Gilchrist, G. (2018). Iffy quotient: A platform health metric for misinformation.

Rivers, C. M. and Lewis, B. L. (2014). Ethical research standards in a world of big data. *F1000Research*.

Rye, E., Blackburn, J., and Beverly, R. (2020). Reading In-Between the Lines: An Analysis of Dissenter. In *ACM IMC*.

Saeed, M. H., Ali, S., Blackburn, J., De Cristofaro, E., Zannettou, S., and Stringhini, G. (2022). Trollmagnifier: Detecting state-sponsored troll accounts on reddit. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 2161–2175. IEEE.

Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S. D., and Vallina-Rodriguez, N. (2018). A long way to the top: significance, structure, and stability of internet top lists. In *ACM IMC*.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21.

Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.

Shahi, G. K., Dirkson, A., and Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on twitter. *Online social networks and media*, 22:100104.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications*, 9(1).

Shu, K., Cui, L., Wang, S., Lee, D., and Liu, H. (2019). defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1).

Soni, S., Ramirez, S. L., and Eisenstein, J. J. (2019). Detecting Social Influence in Event Cascades by Comparing Discriminative Rankers. In *SIGKDD Workshop on Causal Discovery*.

Soroka, S., Fournier, P., and Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *PNAS*, 116(38).

spaCy (2019). Industrial-Strength Natural Language Processing. `https://spacy.io/`.

spaCy (2019). Named Entity Recognition.

spacy (2022). `https://spacy.io/`.

Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. In *ICWSM*.

Stringhini, G., Mourlanne, P., Jacob, G., Egele, M., Kruegel, C., and Vigna, G. (2015). Evilcohort: Detecting communities of malicious accounts on online services. In *24th USENIX Security Symposium (USENIX Security 15)*.

Tahmasbi, F., Schild, L., Ling, C., Blackburn, J., Stringhini, G., Zhang, Y., and Zannettou, S. (2021). go eat a bat, chang!: On the emergence of sinophobic behavior on web communities in the face of covid-19. In *Proceedings of the Web Conference 2021*, pages 1122–1133.

Tasnim, S., Hossain, M. M., and Mazumder, H. (2020). Impact of rumors and misinformation on covid-19 in social media. *Journal of preventive medicine and public health*, 53(3):171–174.

Taulé, M., Martí, M. A., Rangel, F. M., Rosso, P., Bosco, C., Patti, V., et al. (2017). Overview of the task on stance and gender detection in tweets on catalan independence at ibereval 2017. In *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017*, volume 1881, pages 157–177. CEUR-WS.

The Computational Event Data System (2014). Dictionaries. `http://eventdata.parusanalytics.com/software.dir/dictionaries.html`.

The Daily Caller (2016). SOURCES: China Hacked Hillary Clinton's Private E-mail Server. `https://dailycaller.com/2018/08/27/china-hacked-clinton-server`.

The New York Times (2017). Judge Blocks Trump Order on Refugees Amid Chaos and Outcry Worldwide. `https://www.nytimes.com/2017/01/28/us/refugees-detained-at-us-airports-prompting-legal-challenges-to-trumps-immigration-order.html`.

The New York Times (2020). Reddit, acting against hate speech, bans 'the_donald' subreddit. `https://www.nytimes.com/2020/06/29/technology/reddit-hate-speech.html`.

The Washington Post (2016). Trump supporter charged with voting twice in Iowa. `https://www.washingtonpost.com/news/post-nation/wp/2016/10/29/trump-supporter-charged-with-voting-twice-in-iowa/`.

Times, N. Y. (2017). "resist" is a battle cry, but what does it mean? `https://www.nytimes.com/2017/02/14/us/politics/resist-anti-trump-protest.html`.

Trevett, B. (2021). Pytorch sentiment analysis. `https://github.com/bentrevett/pytorch-sentiment-analysis`.

Ullah, A., Das, A., Das, A., Kabir, M. A., and Shu, K. (2021). A survey of covid-19 misinformation: Datasets, detection techniques and open issues. *arXiv preprint arXiv:2110.00737*.

Uscinski, J. E., Enders, A. M., Klofstad, C., Seelig, M., Funchion, J., Everett, C., Wuchty, S., Premaratne, K., and Murthi, M. (2020). Why do people believe covid-19 conspiracy theories? *Harvard Kennedy School Misinformation Review*, 1(3).

Vamvas, J. and Sennrich, R. (2020). X-stance: A multilingual multi-target dataset for stance detection. *arXiv preprint arXiv:2003.08385*.

Van Hoozer, S. and Peuchaud, S. (2020). "Speaking of Sexual Harassers Who Should Resign Tomorrow... Donald Trump": A Feminist Rhetorical Analysis of Stephen Colbert's Late Show Monologues. *The Journal of Popular Culture*, 53(1):34–57.

van Prooijen, J.-W. and Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European journal of social psychology*, 48(7):897–908.

VirusTotal (2020). VirusTotal. https://www.virustotal.com. Accessed: 2021-04-08.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380).

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. (2018). Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.

Wang, J. and Paschalidis, I. C. (2016). Botnet detection based on anomaly and community detection. *IEEE Transactions on Control of Network Systems*, 4(2):392–404.

Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *arXiv:1705.00648*.

Wang, Y., Tahmasbi, F., Blackburn, J., Bradlyn, B., De Cristofaro, E., Magerman, D., Zannettou, S., and Stringhini, G. (2021). Understanding the use of fauxtography on social media. In *International Conference on Web and Social Media*.

Weinzierl, M., Hopfer, S., and Harabagiu, S. M. (2021). Misinformation adoption or rejection in the era of covid-19. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), AAAI Press*.

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Sameer Pradhan, L. R., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., and Houston, A. (2019). OntoNotes Release 5.0. https://catalog.ldc.upenn.edu/LDC2013T19.

Welbers, K. (2016). *Gatekeeping in the Digital Age*. PhD thesis.

WHO (2020). https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-food-safety-and-nutrition.

WHO (2021). https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-hydroxychloroquine.

Wikipedia (2021). https://en.wikipedia.org/wiki/Inverted_pyramid_(journalism).

Wikipedia (2022). https://en.wikipedia.org/wiki/Conspiracy_theory.

Wilson, T. and Starbird, K. (2020). Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review*, 1(1).

Wilson, T., Zhou, K., and Starbird, K. (2018). Assembling strategic narratives: Information operations as collaborative work within an online community. In *CSCW*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Wong, J. C. (2018). What is qanon? explaining the bizarre rightwing conspiracy theory. *The Guardian*.

Wray, M. (2020). corona challenge: Tiktok star films herself licking airplane toilet seat. `https://globalnews.ca/news/6718358/tiktok-toilet-seat-lick-coronavirus/`.

Wu, L. and Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *WSDM*.

Wu, L., Morstatter, F., Carley, K. M., and Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, 21(2):80–90.

Zannettou, S. (2021). " i won the election!": An empirical analysis of soft moderation interventions on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 865–876.

Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., and Blackburn, J. (2018a). What is gab: A bastion of free speech or an alt-right echo chamber. In *The WebConf Companion*.

Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., and Suarez-Tangil, G. (2018b). On the origins of memes by means of fringe web communities. In *ACM IMC*.

Zannettou, S., Caulfield, T., Bradlyn, B., De Cristofaro, E., Stringhini, G., and Blackburn, J. (2020a). Characterizing the use of images in state-sponsored information warfare operations by russian trolls on twitter. In *ICSWM*.

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Sirivianos, M., Stringhini, G., and Blackburn, J. (2017). The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. In *ACM IMC*.

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., and Blackburn, J. (2019a). Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *The WebConf Companion*.

Zannettou, S., Caulfield, T., Setzer, W., Sirivianos, M., Stringhini, G., and Blackburn, J. (2019b). Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls. In *WebSci*.

Zannettou, S., Finkelstein, J., Bradlyn, B., and Blackburn, J. (2020b). A quantitative approach to understanding online antisemitism. In *ICWSM*.

Zhang, D. Y., Shang, L., Geng, B., Lai, S., Li, K., Zhu, H., Amin, M. T., and Wang, D. (2018). Fauxbuster: A content-free fauxtography detector using social media comments. In *IEEE Big Data*.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., and Li, X. (2011). Comparing twitter and traditional media using topic models. In *European Conference on Information Retrieval*.

Zheng, X., Han, J., and Sun, A. (2018). A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1652–1671.

Zhou, K., Zha, H., and Song, L. (2013). Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Artificial Intelligence and Statistics*.

Zhou, X., Mulay, A., Ferrara, E., and Zafarani, R. (2020). Recovery: A multimodal repository for covid-19 news credibility research. *arXiv:2006.05557*.

Zhou, X. and Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.

Ziems, C., He, B., Soni, S., and Kumar, S. (2020). Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. *arXiv preprint arXiv:2005.12423*.

Zillmann, D., Gibson, R., and Sargent, S. L. (1999). Effects of photographs in news-magazine reports on issue perception. *Media Psychology*, 1(3).

Zillmann, D., Knobloch, S., and Yu, H.-s. (2001). Effects of photographs on the selective reading of news reports. *Media Psychology*, 3(4).

Zlatkova, D., Nakov, P., and Koychev, I. (2019). Fact-checking meets fauxtography: Verifying claims about images. In *EMNLP-IJCNLP*.