CoMID: COVID-19 Misinformation Alignment Detection Using Content and User Data

Nazanin Jafari,[1] **Sheikh Muhammad Sarwar,** [2] **James Allan** [1] **Keen Sung,**[3] **Shiri Dori Hacohen,**[3] **Matthew Rattigan,**[1]

[1] University of Massachusetts Amherst
[2] Amazon
[2] AuCoDe
{nazaninjafar,allan,rattigan}@cs.umass.edu, smsarwar@amazon.com, {keen,shiri}@aucode.io

Abstract

An important approach to understanding and mitigating the spread of misinformation is to recognize whether a given social media post aligns with the false information (that is, *agreeing* with it) or *disagreeing* with it. In this paper, we present CoMID, a method that detects whether a tweet agrees or disagrees with a misinformation claim, based on the content of the tweet and the author's propensity to spread misinformation. To calculate the propensity of the user, we utilize their past tweets and profile description based on a new model. We evaluate this method on our newly introduced dataset, "COVID-Myths", and compare it to existing state-of-the-art content-only and user & content-based methods. In general,the proposed model, CoMID, is beneficial and achieves a 5% performance gain (in terms of the F1 score) compared to the best-performing baseline. Additionally, we evaluate the generalizability of CoMID in a zero-shot setting by leveraging only the weakly supervised data. CoMID achieves state-of-the-art performance in this setting, which suggests the effectiveness of utilizing user data to capture propensity.

Introduction

The wide dissemination of misinformation on social media has detrimental effects on society. It surfaces on different online platforms, and different topics such as finance, politics, health, etc. Since early 2020, the world has encountered a wave of misinformation around the COVID-19 pandemic, which posed a profound threat to public health.

The most common approaches to address misinformation in general are post-based methods using textual content and using lexical and syntactic features of the post (e.g. (Ma et al. 2016; Wang et al. 2021; Silva et al. 2021)). Deep learning-based methods also have recently shown promising results in this area (e.g., (Ruchansky, Seo, and Liu 2017; Zhang et al. 2020; Vlad et al. 2019; Ding, Hu, and Chang 2020; Weinzierl, Hopfer, and Harabagiu 2021)). Another line of studies combines knowledge sources with posts to improve predictions (e.g., (Sheng et al. 2022; Cui et al. 2020; Hu et al. 2021; Pan et al. 2018; Dun et al. 2021; Thorne et al. 2018; Aly et al. 2021; Shaar et al. 2020)) which rely only on the content features of the post and might be domain dependent and less likely to generalize to other domains or topics.

We formulate the detection of misinformation as a *misinformation alignment* problem. Given social media posts, our task is to develop a model to predict the stance of a given social media post towards the pre-defined set of false claims. We utilize user information. Users have valuable information that can help detect misinformation alignment (Shu et al. 2019). An important source of useful features are their historical posts which have been helpful in detecting fake news (Dou et al. 2021). We use historical user posts and explore this problem in a real-world scenario, where a new false claim and its relevant social media posts are not available. In our model CoMID, we first retrieve the user posts that are most similar to the candidate misinformation post (which we call the target post/tweet from now on), then we capture the user's propensities to spread misinformation using these posts.

To evaluate our model, we introduce the COVID myths dataset, with the alignment of social media posts to previously debunked COVID-19 myths. This dataset consists of 8 claims that have emerged and been debunked during the pandemic by verified sources and $3,528$ labeled tweets. These tweets have been annotated as either "agreeing with a misinformation claim, "disagreeing with it, or "neither".

Additionally, we evaluated our model's ability to incorporate new claims by transfer learning on four publicly available COVID-19 misinformation datasets. We pre-train our model, CoMID on this large weakly labeled dataset and evaluate it on COVID-Myths in zero-shot and limited-data settings (i.e., we used a few hundred data cases from COVID-Myths for training). Our findings show that our model can successfully use out-of-domain data to improve performance on the COVID-Myths dataset. Our main contributions can be summarized as follows:

- We introduce CoMID which utilizes the historical posts of the users to capture the propensity to misinformation in a novel approach.
- We introduce a new COVID-Myths dataset with 8 false claims about COVID-19 with $3,528$ labeled tweets.
- Experimental analysis of unseen false claims shows that CoMID has 5% performance gain in F1 score over the state-of-the-art model (Dou et al. 2021). Furthermore, we show that CoMID benefits greatly from transfer learning on in-domain weakly supervised dataset compared to the

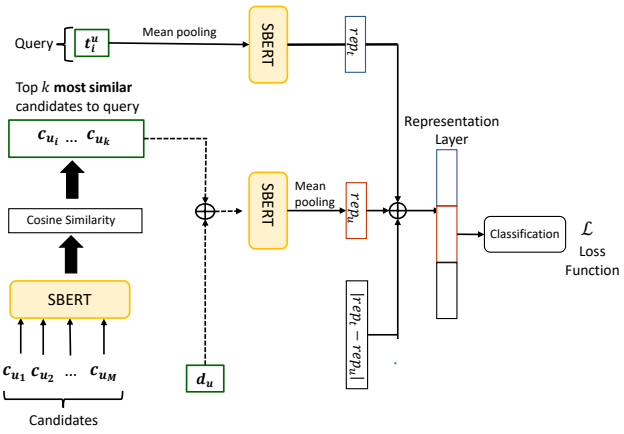

Figure 1: Overview of the proposed CoMID framework. Given a tweet, (1) we extract sentence embedding rep_t, and (2) we concatenate the top k tweets of the user, with user description d_u and extract their sentence embeddings rep_u and concatenate these representations and their absolute difference $|rep_t - rep_u|$ in a representation layer to predict misinformation alignment.

best-performing SOTA model, which suggests the effectiveness of CoMID in detecting misinformation on never or rarely seen topics.

Framework

We detect the misinformation alignment of a social media post with one or more claims in our database of false claims (i.e., *agreeing*). A non-alignment indicates that the post is either irrelevant or neutral or refutes the misinformation and/or provides counter arguments (i.e., *disagreeing, neither*). We utilize social media posts from Twitter, and thus we define the posts as tweets from now on.

We have a dataset, $T_L = \{t_1^1, t_2^2, \ldots, t_N^u\}$ of N tweets. Here, t_i^u is a tweet from user u and $\mathcal{S}_u = \{c_{u_1}, c_{u_2}, \ldots, c_{u_M}\}$ is the set of historical tweets from u. Here, c_{u_i} is the i-th tweet from the timeline of u. T_L and $\mathcal{S}_u$ are disjoint, which means that the dataset of tweets and historical tweets are separate. In addition, we have the profile description d_u, which is a sequence of tokens representing u. We use y_i to define the misinformation alignment label of tweet t_i^u, which can be *agree*, *disagree*, or *neither*. The label *agree* indicates that a tweet aligns with a false claim from a set of pre-defined ones, whereas *disagree* indicates misalignment of a tweet with misinformation and *neither* corresponds to tweets that are either unrelated or pose no stance towards the misinformation. Thus, our data instance is a tuple of $(t_i^u, \mathcal{S}_u, d_u)$, for which our task is to predict the false claim alignment y_i of t_i^u. Thus, unlike existing text classification approaches that rely on encoding textual content, we also encode the user profile based on description and history for effective alignment detection.

As we focus on representing both content and user, our model has two main components: 1) *Content encoder*, and 2) *User encoder*. The former extracts textual information from t_i^u by obtaining its vectorized representation using pre-trained language models. In particular, we use a pretrained SBERT (Reimers and Gurevych 2019) to encode t_i^u. The pre-trained encoder returns the representation of the tokens in t_i^u. Finally, we apply the mean pooling to the token sequence to obtain an aggregated representation, rep_t.

The *User encoder* captures the propensity of users to spread misinformation by representing their historical tweets, $\mathcal{S}_u$ and their profile description d_u. Not all the past tweets of the user are related to the target tweet. Thus, to encode user information for a proclivity to mis/information, we employ only the most relevant tweets of the user to the target user. This is achieved by setting the target tweet t_i^u as a query and the set of tweets in $\mathcal{S}_u$ as candidates. We use SBERT representation with cosine similarity to compute the semantic textual similarity between the query and each of the candidates. We select the top$k$ similar candidates to represent the user. This retrieval approach benefits the model both in revealing the most relevant tweets and in filtering out many user tweets that do not provide useful information. Another important user feature is their profile description, d_u, which provides valuable information about the user and could provide a summary of their timeline tweets. We also utilize this information in our user encoder.

Given k past tweets (that is,$\{c_{u1}, .., c_{uk}\}$ and the description of the user profile d_u, we concatenate these inputs separated by the [SEP] token and encode the concatenated sequence using SBERT. We concatenate these user tweets as a single sequence. Longer sequences are truncated to the SBERT sequence-length limitation (512 tokens). Finally, to obtain user representation, similar to the content encoder, we apply mean pooling to this sequence and name it rep_u.

Finally, we incorporate the output of each of the encoders above (i.e., rep_t and rep_u) by concatenation in a final representation layer. This representation layer is then fed into a neural classifier with a fully connected layer to predict misinformation alignment. We then optimize cross-entropy loss. Figure 1 shows the details of the CoMID model.

Dataset Construction

Our dataset, COVID-Myths, is a collection of COVID-19-related tweets that we annotate with false claims. We use eight fact-checked false claims[1], factcheck.org[2], PolitiFact[3] and obtain an annotation for a tweet as the entitlement between the tweet and a false claim.

We use Twitter's search API to obtain tweets that match a list of COVID-19-related terms and that were published between January 2020 and April 2021. Then, for each of the eight false claims, we compute its similarity score with each of the collected tweets. The similarity score is based on the cosine similarity of the SBERT representations of the false claim and a tweet. For each false claim, we consider the top 500 similar tweets for human annotation. We obtain annotations from Mechanical Turk [4] annotator, who determines the entitlement of each of the tweets toward the false claim

[1]https://www.snopes.com/
[2]https://www.factcheck.org/
[3]https://www.politifact.com/
[4]https://www.mturk.com/

query. We use "agree", "disagree" or "neither" as labels to obtain judgments for entitlement.

We carry out the annotation task based on an approved IRB protocol. Annotators were paid according to the minimum wage of the MA, United States. Annotators are located in the United States and are at least 18 years of age. We use an averaging metric for inter-annotator agreement and compute it by averaging over majority votes over all the votes. Finally, we evaluate the quality of the annotations and validate the data measures of inter-annotator agreement. We have an 85.8% inter-annotator agreement from our annotators. To evaluate the quality of the annotations, we randomly selected fifteen tweets from each false claim (5 from each previously labeled class) and asked workers (who are different than the previous workers) on Mechanical Turk to re-annotate them. Overall, we had an agreement of 85.8% between the annotators. Our annotated $3,528$ tweets can be broken down into 1248 *agree* labeled tweets, 1082 *disagree*, and 1199 *neither* belonging to 3021 unique users. We will release this dataset for public use.

Experiments

We use PyTorch[5] and Huggingface libraries (Wolf et al. 2020) to implement our models. We train our model on an NVIDIA GeForce RTX 2080 Ti device with a learning rate of $2e-5$ and a batch size of 8. In order to find the best value for the number of similar historical tweets to represent a user, we experimented with five, ten, and twenty similar tweets. We found ten similar tweets to be optimal based on the performance of the validation set.

To take advantage of the past tweets of a user for computing the user representation, we concatenate them to form a single input. Our data analysis reveals that the mean of the sequence lengths of the historical tweets from users is 23 tokens. Thus, we set the maximum length of each of the ten historical tweets at 40, resulting in a concatenated sequence that is 400 tokens long. We also set the maximum sequence length of the tweet we classify at 200 in all experiments.

We report F1, Precision, Recall and Accuracy after macro averaging over three classes of *agree*, *disagree*, and *neither*. All metrics are averaged over 5 runs initialized with five random seeds. We compare the performance of CoMID against several state-of-the-art baseline models that consider content and/or user representations for misinformation alignment detection. The baselines are described below:

UPFD-EXO (Dou et al. 2021) has two encoders: i) an endogenous encoder to encode content and historical posts; and ii) an exogenous encoder which is a graph neural network of users. As we do not have user connectivity information in our dataset, we do not utilize any graph networks. We implement UPFD without an exogeneous encoder named UPFD-EXO. In this model, each historical post of the user is given as a sequence to BERT and then averaged over all user posts to obtain a single-user presentation and the model does not have a dynamic user encoder to be fine-tuned along with a content encoder. Note that UPFD-EXO utilizes Bert-Large-Cased for averaging historical posts as embedding. Since in our study we use SBERT, for a fair comparison, we also implemented UPFD-EXO using SBERT as an encoder, and we name it UPFD-EXO (SBERT). We additionally evaluate UPFD-EXO (SBERT) in the $k = 10$ most similar user tweets setting and name it UPFD-EXO$_{10}$ (SBERT).

D-AVG stands for Dynamic Averaging. After extracting k most similar user tweets, instead of concatenating these tweets, we encoded each tweet using SBERT separately. The output of this encoder is a $k \times d$ dimensional vector where d refers to the dimension of each user tweet representation. Finally, we averaged these k vectors to get a single-user representation. Unlike UPFD-EXO, in this encoder, the user representation is not computed beforehand, i.e., the user encoder is fine-tuned along with the content encoder in the training phase to leverage content information.

I-XLM-R (Kazemi et al. 2021) This model is a multilingual embedding model trained for claim-matching the previously fact-checked claims.

RoBERTa-Base is the base cased RoBERTa (Liu et al. 2019).

SBERT is sentence BERT. We fine-tune these models with our data and report results on the test dataset.

Results

Generalizability to Unseen Claims. We evaluate the effectiveness of our model in a real-world scenario where we might receive social media posts in a stream related to a wide range of emerging claims. In other words, we train on data belonging to 7 out of the 8 false claims in our dataset (COVID-Myths) and leave data belonging to one of the false claims as test data only.

The experimental results for this scenario are shown in Table 1. Our proposed method achieves a performance gain of around 5% in both the F1 score and the accuracy compared to the best-performing baseline, UPFD-EXO (SBERT). A comparison between UPFD-EXO (SBERT) and UPFD-EXO also suggests that using a sentence-based BERT encoder achieves better performance. Additionally, our modification of UPFD-EXO (SBERT) to average over only 10 most similar user tweets UPFD-EXO$_{10}$ (SBERT) instead of all user tweets shows a minimal performance loss which suggests a nontrivial advantage for using all user posts.

In general, these results suggest that CoMID is able to mitigate the difficulty of predicting tweets about an unseen claim better than the baselines. Moreover, it shows the robustness and generalizability of CoMID in scenarios that are closer to misinformation detection in the wild where social media posts can be received in a stream of new claims.

Pre-training on Weakly labeled data. We examine the effectiveness of our method in a larger domain. To this end, we train our model and the baselines on a combined collection of two human-labeled and two weakly labeled datasets. We evaluated the trained models in our COVID-Myths dataset to investigate how transferable COVID-19

[5]https://pytorch.org/

Table 1: Comparison between different baseline models. The results are given as the average of 8 runs, leaving one false claim out for testing. Statistically significant results (based on the t-test) of the best-performing baselines (UPFD-EXO (SBERT)) compared to CoMID are denoted by a star ($*p \leq 0.01$) and the best scores are presented in bold font.

Model	$F1_{macro}$	Acc	Agree			Disagree			Neither		
			F1	P	R	F1	P	R	F1	P	R
UPFD-EXO	44.9	48.9	29.3	57.9	22.2	47.7	62.4	43.3	55.0	43.5	77.4
UPFD-EXO (SBERT)	49.1	52.8	34.4	64.9	23.5	56.6	65.2	50.0	56.3	45.1	75.1
$UPFD\text{-}EXO_{10}$ (SBERT)	50.0*	53.0*	34.8*	66.6*	26.4*	58.6*	65.5*	57.2*	56.7*	46.5*	75.6
D-AVG	49.9	52.7	36.7	63.0	29.0	56.3	64.5	55.9	56.7	47.7	73.5
RoBERTa	48.2	51.1	32.6	63.8	24.3	55.6	66.4	52.9	56.4	45.8	76.5
SBERT	49.3	51.4	34.1	64.2	25.1	57.2	65.1	54.3	56.5	46.0	74.9
I-XLM-R	43.2	47.5	26.0	59.1	18.9	48.8	59.6	46.4	54.8	43.8	76.1
CoMID	**51.0***	**53.8***	**36.4***	**65.4***	**28.2***	**59.6***	**65.9***	**58.4***	**56.9***	**47.3***	**74.9***

misinformation models are. The datasets we use for training are COVIDLIES (Hossain et al. 2020), COVID-CQ (Mutlu et al. 2020), COAID (Cui and Lee 2020), and ReCOVery (Zhou et al. 2020). COVIDLIES (Hossain et al. 2020) consists of 6761 expert-annotated tweets related to 86 misconceptions on whether they agree, disagree, or have no stance towards them. COVID-CQ consists of more than 14 thousand annotated tweets on "the use of chloroquine for COVID-19 treatment." topic Agree/Disagree and No stance labels. CoAID (Cui and Lee 2020) is a weakly-labeled dataset that is derived from the engagement of 296000 users with 4251 fake/real labeled news articles. ReCOVery (Zhou et al. 2020) is derived from 140000 social engagements with more than 2000 reliable/unreliable news articles. To align labels, we use the Twitter engagement data from the CoAID and ReCOVery datasets. Given fake/unreliable news/claims and the corresponding engagement tweets, we pair them as premise and hypothesis and input them into the RoBERTa-MNLI model and the output is used as weak labels.

we pre-train CoMID on this collection of more than 70K weakly labeled datasets and use this model for transfer-learning on the COVID-Myths dataset. We evaluate this model in two settings: 1) In the "zero-shot" setting where we do not fine-tune any COVID-Myths dataset and 2) In the "limited-data" setting where we fine-tune a few randomly sampled "COVID-Myths" instances. We pre-train both CoMID and UPFD-EXO and compare their performance. Figure 2 shows the results with respect to this experiment. As can be seen, CoMID can leverage weakly labeled dataset better than UPFD-EXO (SBERT) both in the limited-data setting and the zero-shot setting. Moreover, we observe that CoMID is able to close the performance gap between full-training and limited-training faster and with fewer data compared to UPFD-EXO (SBERT) indicating better domain adaptation and generalizability of the proposed method. However, the substantial performance gap between zero-shot setting and full training on the COVID-Myths dataset both in CoMID and UPFD-EXO shows the need for improvement in such approaches.

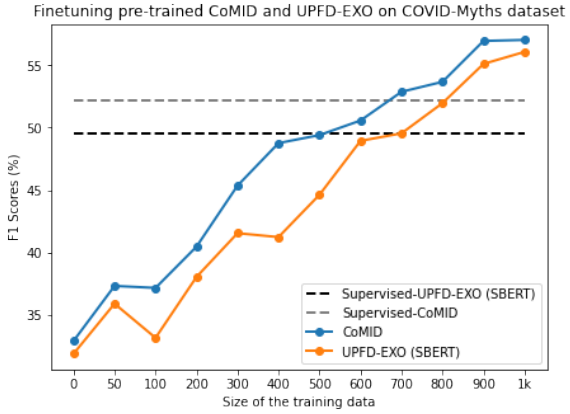

Figure 2: Comparison between pre-trained CoMID and UPFD-EXO (SBERT) in zero-shot setting (with 0 added data from COVID-Myths for fine-tuning) and limited-data setting.

Conclusion

We propose CoMID and incorporate user timeline posts along with user bios to capture the misinformation propensity of users for COVID-19 misinformation alignment detection on Twitter. Our extensive evaluations show that CoMID outperforms the state of the art in the leave-one-claim-out setting. We also show that pre-training this method on weakly labeled dataset facilitate transfer learning by needing fewer training data for fine-tuning. Furthermore, we introduce the "COVID-Myths" dataset with 3528 annotated tweets.

Acknowledgments

This work was supported in part by the Center for Intelligent Information Retrieval and in part by NSF grant number 1813662. Early work in this project was supported by NSF grant numbers 1951091 and 2147305. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

References

Aly, R.; Guo, Z.; Schlichtkrull, M. S.; Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; Cocarascu, O.; and Mittal, A. 2021. The Fact Extraction and VERification Over Unstructured and Structured information (FEVEROUS) Shared Task. In *Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER)*, 1–13. Dominican Republic: Association for Computational Linguistics.

Cui, L.; and Lee, D. 2020. CoAID: COVID-19 Healthcare Misinformation Dataset. arXiv:2006.00885.

Cui, L.; Seo, H.; Tabar, M.; Ma, F.; Wang, S.; and Lee, D. 2020. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 492–502.

Ding, J.; Hu, Y.; and Chang, H. 2020. BERT-Based Mental Model, a Better Fake News Detector. In *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 396–400.

Dou, Y.; Shu, K.; Xia, C.; Yu, P. S.; and Sun, L. 2021. User Preference-Aware Fake News Detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, 2051–2055. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380379.

Dun, Y.; Tu, K.; Chen, C.; Hou, C.; and Yuan, X. 2021. Kan: Knowledge-aware attention network for fake news detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 81–89.

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*.

Hu, L.; Yang, T.; Zhang, L.; Zhong, W.; Tang, D.; Shi, C.; Duan, N.; and Zhou, M. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. 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: Long Papers)*, 754–763.

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*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B. J.; Wong, K.-F.; and Cha, M. 2016. Detecting rumors from microblogs with recurrent neural networks.

Mutlu, E. C.; Oghaz, T.; Jasser, J.; Tutunculer, E.; Rajabi, A.; Tayebi, A.; Ozmen, O.; and Garibay, I. 2020. A stance data set on polarized conversations on Twitter about the efficacy of hydroxychloroquine as a treatment for COVID-19. *Data in brief*, 106401.

Pan, J. Z.; Pavlova, S.; Li, C.; Li, N.; Li, Y.; and Liu, J. 2018. Content based fake news detection using knowledge graphs. In *International semantic web conference*, 669–683. Springer.

Reimers, N.; and Gurevych, I. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Ruchansky, N.; Seo, S.; and Liu, Y. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797–806.

Shaar, S.; Babulkov, N.; Da San Martino, G.; and Nakov, P. 2020. That is a Known Lie: Detecting Previously Fact-Checked Claims. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3607–3618. Online: Association for Computational Linguistics.

Sheng, Q.; Cao, J.; Zhang, X.; Li, R.; Wang, D.; and Zhu, Y. 2022. Zoom Out and Observe: News Environment Perception for Fake News Detection. *arXiv preprint arXiv:2203.10885*.

Shu, K.; Zhou, X.; Wang, S.; Zafarani, R.; and Liu, H. 2019. The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 436–439.

Silva, A.; Luo, L.; Karunasekera, S.; and Leckie, C. 2021. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 557–565.

Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; and Mittal, A. 2018. FEVER: a Large-scale Dataset for Fact Extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 809–819. New Orleans, Louisiana: Association for Computational Linguistics.

Vlad, G.-A.; Tanase, M.-A.; Onose, C.; and Cercel, D.-C. 2019. Sentence-Level Propaganda Detection in News Articles with Transfer Learning and BERT-BiLSTM-Capsule Model. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, 148–154.

Wang, Y.; Ma, F.; Wang, H.; Jha, K.; and Gao, J. 2021. Multimodal emergent fake news detection via meta neural process networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3708–3716.

Weinzierl, M. A.; Hopfer, S.; and Harabagiu, S. M. 2021. Misinformation Adoption or Rejection in the Era of COVID-19. In *ICWSM*, 787–795.

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*, 38–45. Online: Association for Computational Linguistics.

Zhang, T.; Wang, D.; Chen, H.; Zeng, Z.; Guo, W.; Miao, C.; and Cui, L. 2020. BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Zhou, X.; Mulay, A.; Ferrara, E.; and Zafarani, R. 2020. Recovery: A multimodal repository for covid-19 news credibility research. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 3205–3212.