

On Unifying Misinformation Detection

Nayeon Lee^{§*} Belinda Z. Li^{†‡} Sinong Wang[†]
Pascale Fung[§] Hao Ma[†] Wen-tau Yih[†] Madian Khabisa[†]

[§]Hong Kong University of Science and Technology

[†]Facebook AI

[‡]MIT CSAIL

nyleeaa@connect.ust.hk, mkhabisa@fb.com

Abstract

In this paper, we introduce UNIFIEDM2, a general-purpose misinformation model that jointly models multiple domains of misinformation with a single, unified setup. The model is trained to handle four tasks: detecting *news bias*, *clickbait*, *fake news* and verifying *rumors*. By grouping these tasks together, UNIFIEDM2 learns a richer representation of misinformation, which leads to state-of-the-art or comparable performance across all tasks. Furthermore, we demonstrate that UNIFIEDM2’s learned representation is helpful for few-shot learning of unseen misinformation tasks/datasets and model’s generalizability to unseen events.

1 Introduction

On any given day, 2.5 quintillion bytes of information are created on the Internet, a figure that is only expected to increase in the coming years (Marr, 2018). The internet has allowed information to spread rapidly, and studies have found that misinformation spreads quicker and more broadly than true information (Vosoughi et al., 2018). It is thus paramount for misinformation detection approaches to be able to adapt to new, emerging problems in real time, without waiting for thousands of training examples to be collected. In other words, the *generalizability* of such systems is essential.

Misinformation detection is not well-studied from a generalizability standpoint. Misinformation can manifest in different forms and domains, i.e., fake news, clickbait, and false rumors, and previous literature has mostly focused on building specialized models for a single domain (Rubin et al., 2016; Omidvar et al., 2018; Ma et al., 2018). (Even prior literature on multi-tasking for misinformation (Kochkina et al., 2018) focuses more on

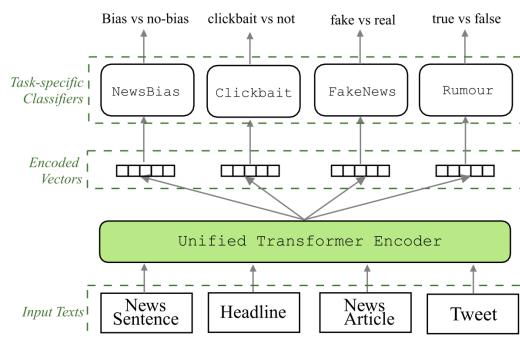


Figure 1: Architecture of our UNIFIEDM2 model

using auxiliary tasks to boost performance on a *single* task, rather than on all tasks.) However, though these domains may differ in format (long articles vs. short headlines and tweets) and exact objective (“is this fake” vs. “is this clickbait”), they have the same ultimate goal of deceiving their readers. As a result, their content often exhibits similar linguistic characteristics, such as using a sensational style to incite curiosity or strong emotional responses from readers. Furthermore, models trained on multiple tasks are more robust and less prone to overfitting to spurious domain-specific correlations. Thus, unifying various domains of misinformation allows us to build a generalizable model that performs well across multiple domains/formats of misinformation.

In this work, we propose Unified Misinfo Model (UNIFIEDM2), a misinformation detection model that uses multi-task learning (Caruana, 1997; Maurer et al., 2016; Zhang and Yang, 2017) to train on different domains of misinformation. Through a comprehensive series of empirical evaluations, we demonstrate that our approach is effective on *all* tasks that we train on, improving F_1 in some cases by an absolute $\sim 8\%$. Moreover, we conduct ablation studies to more precisely characterize how such positive transfer[†] is attained. Beyond improvements on seen datasets, we examine the gen-

*Work partially done while interning at Facebook AI.

[†]Work partially done while working at Facebook AI.

Task	Dataset Name	Granularity	Labels (Positive/Negative)	Dataset Size	Positive Class Size
NEWSBIAS	BASIL	sentence	contains-bias/no-bias	7,984	1,727
FAKENEWS	Webis	article	fake/true	1,627	363
RUMOR	PHEME	tweet	fake/true	1,705	1,067
CLICKBAIT	Clickbait	headline	is-clickbait/not-clickbait	19,538	4,761

Table 1: Summary of the four misinformation datasets we train on with UNIFIEDM2.

eralizability of our proposed approach to unseen tasks/datasets and events. This is highly applicable to real-world use cases, where obtaining new misinformation labels is costly and systems often wish to take down misinformation in real time. Our experimental results indicate that our unified representation has better generalization ability over other baselines.

2 UNIFIEDM2

In this section, we describe the architecture and the training details for our proposed UNIFIEDM2 model.

2.1 Architecture

Our proposed model architecture is a hard-parameter sharing multi-task learning model (Ruder, 2017), where a single shared RoBERTa (Liu et al., 2019b) encoder is used across all tasks. RoBERTa is a Transformer encoder pretrained with a masked-language-modeling objective on English Wikipedia and news articles (CC-NEWS), among other data. We additionally append task-specific multi-layer perceptron (MLP) classification heads following the shared encoder. During multi-task training, the model sees examples from all datasets, and we jointly train the shared encoder with *all* task-specific heads. During inference time, we only use the classification head relevant to the inference-time task. The overall architecture of the model is shown in Figure 1.

2.2 Training

Our model training process consists of two steps. The first step is multi-task training of the shared UNIFIEDM2 encoder to learn a general misinformation representation. We jointly optimize for all tasks $t_1 \dots t_T$ by optimizing the sum of their task-specific losses L_t , where L_t refers to the cross-entropy loss of the task-specific MLP classifiers. Our overall loss is defined as $L_{\text{multi}} = \sum_{t=t_1 \dots t_T} L_t$. Note that since the dataset sizes

are different, we over-sample from the smaller datasets to make the training examples roughly equal. The second step is to fine-tune each task-specific heads again, similarly to the MT-DNN by Liu et al. (2019a), to obtain the results reported in Table 2 and Table 4.

3 Experiment

Here, we provide experimental details (dataset, baselines, experimental setups) and results that empirically show the success of the proposed UNIFIEDM2 model.

3.1 Misinformation Tasks/Dataset

Table 1 lists the four misinformation tasks/datasets we use to train UNIFIEDM2. They span various granularities and domains (articles, sentences, headlines and tweets) as well as various objectives (classifying veracity, bias and clickbaity-ness).

NEWSBIAS A task to classify whether a given sentence from a news article contains political bias or not. We adapt the BASIL (Fan et al., 2019) dataset, which has bias-span annotations for lexical and informational bias within news articles. Using this dataset, we also include two auxiliary tasks related to political-bias detection: 1) *bias type classification* – given a biased sentence, the type of the bias (lexical vs informational) is classified; and 2) *polarity detection* – given a biased sentence, its polarity (positive, negative, neutral) is determined.

FAKENEWS An article-level fake news detection task that leverages the Webis (Potthast et al., 2018) dataset annotated by professional journalists.

RUMOR A task to verify the veracity of a rumor tweet. The PHEME dataset (Zubiaga et al., 2016), which contains rumor tweets with their corresponding reply tweets (social engagement data), is used for this task. We only use the text of the source rumor tweet since we focus on learning a good representation for misinformation *text*. Originally,

Tasks	SoTA models		RoBERTa		UNIFIEDM2	
	Acc	F1	Acc	F1	Acc	F1
NEWSBIAS	N/A	32.0% 43.0%	72.8%	65.5%	81.0%	70.2%
FAKENEWS	58.0%	46.0%	84.3%	74.9%	85.4%	73.9%
RUMOR	81.0%	80.0%	87.6%	86.9%	92.9%	92.5%
CLICKBAIT	83.0%	57.0%	84.4%	77.4%	86.3%	78.7%

Table 2: Results of single-task SoTA papers, the single-task RoBERTa baseline, and our UNIFIEDM2 on all misinformation tasks. SoTA numbers for NEWSBIAS, FAKENEWS and RUMOR are from Fan et al. (2019), Potthast et al. (2018), and Wu et al. (2019), respectively. CLICKBAIT numbers are from running the released code from Omidvar et al. (2018). All the RoBERTa and UNIFIEDM2 results are the averaged results of three seed runs.

there were three class labels (true, false, unverified); however, following other literature (Derczynski et al., 2017; Wu et al., 2019), we report the binary version, excluding the unverified label.

CLICKBAIT A task to detect the clickbaity-ness of news headlines, which refers to sensational headlines that might deceive and mislead readers. For this task, we use the dataset from the Clickbait Challenge.¹

3.2 Baseline Models

State-of-the-Art Models For each misinformation task, we report and compare our approach to the SoTA models from Fan et al. (2019) for NEWSBIAS,² Potthast et al. (2018) for FAKENEWS, Wu et al. (2019) for RUMOR and Omidvar et al. (2018) for CLICKBAIT.

RoBERTa-based Baselines In addition to each task’s published SoTA model, we create RoBERTa-based models by fine-tuning RoBERTa to each individual task.

3.3 Experimental Setup

Training Details We ran all our experiments for 3 times with different shots, and report the average. Our UNIFIEDM2 model is based on RoBERTa-large model which has 355M parameters.

We used the Adam optimizer (Kingma and Ba, 2014) with a mini-batch size of 32. The learning rate was set to 5e-6 with linear learning rate decay. The maximum epoch count was 15, with early stopping patience set to 5. The maximum sequence length of input was set to 128. These parameters

¹<https://www.clickbait-challenge.org/>. There are two versions of the labeled dataset, but we only use the larger one.

²They report bias-detection performance separately on the “lexical-bias vs. no-bias” setting and “informational-bias vs. no-bias” setting. In our experiments, we treat both lexical-bias and informational-bias to be “contains-bias” class, and conduct one unified experiment.

#	Task Combination	Acc	F1
1	RUMOR ST RoBERTa	87.6%	86.9%
	RUMOR, NEWSBIAS	85.9%	85.4%
2	RUMOR, CLICKBAIT	88.8%	87.8%
	RUMOR, FAKENEWS	78.8%	78.7%
3	NEWSBIAS, FAKENEWS, RUMOR	88.2%	87.5%
	NEWSBIAS, RUMOR, CLICKBAIT	91.8%	90.5%
4	FAKENEWS, RUMOR, CLICKBAIT	88.8%	87.8%
	UNIFIEDM2	92.9%	92.5%

Table 3: Ablation study for understanding which task(s), when trained in combination with RUMOR, are most beneficial when evaluated on RUMOR.

were obtained by performing grid-search over our validation loss. We search within the following hyper-parameter bounds: $LR = \{5e - 5, 5e - 6, 5e - 7\}$, $batch = \{16, 32\}$.

Training Details for few-shot experiments We did not do any parameter searching for these few-shot experiments. We kept all the training details and parameters the same to the training details that are state above.

Computing Infrastructure We ran all experiments with 1 NVIDIA TESLA V100 GPU with 32 GB of memory.

3.4 Main Results

Table 2 presents the results of our proposed unified model, UNIFIEDM2, along with the two groups of baseline models. UNIFIEDM2 achieves better or comparable results over both baselines for *all four* misinformation tasks. The improvement is especially prominent on the NEWSBIAS and RUMOR tasks, where we see an 8% and 5% improvement in accuracy, respectively.

3.5 Task Ablation Study

We conduct an ablation study to better understand how other tasks help in our multitask framework.

Tasks	10 examples			25 examples			50 examples		
	Vanilla	ST average	UNIFIEDM2	Vanilla	ST average	UNIFIEDM2	Vanilla	ST average	UNIFIEDM2
Propaganda	45.10%	55.50%	56.19%	56.4%	60.7%	62.5%	56.4%	65.5%	72.9%
Fake News Article	32.97%	38.70%	42.42%	35.0%	58.1%	53.1%	35.0%	67.3%	74.2%
Fake News Title	34.13%	64.08%	55.36%	33.9%	66.6%	67.0%	33.9%	73.5%	71.4%
Covid Check-worthy Twitter	52.17%	55.22%	61.70%	37.3%	62.5%	66.4%	37.3%	65.6%	73.2%
Covid False Twitter Claim	46.93%	48.01%	54.25%	46.4%	51.2%	56.3%	46.4%	54.0%	59.7%
Average	42.26%	52.30%	53.98%	41.80%	59.84%	61.05%	41.80%	65.19%	70.27%

Table 4: Macro-F1 scores of the few-shot experiment with 10, 25, and 50 examples on unseen misinformation-related datasets/tasks. The following datasets are used for each task: PROPAGANDA: Propaganda detection, POLITIFACT: fake news article detection, BUZZFEED: fake news title detection, COVIDTWITTER: Check-worthy twitter detection and false twitter claim detection (this dataset is used for two tasks)

One question we ask is what *kinds* of tasks benefit the most from being trained together. Namely, how well do more “similar” vs. more “different” kinds of task transfer to each other?

Specifically, we use the RUMOR dataset as a case study³. We train on multiple task combinations and evaluate their performance on RUMOR. Results are shown in Table 3. Note that adding FAKE-NEWS alone to single-task RoBERTa, or NEWS-BIAS, actually *hurts* performance, indicating that multi-task learning is not simply a matter of data augmentation. We hypothesize that the drop is due to FAKENEWS being the least similar in format and style to RUMOR. Qualitatively, we compare examples from FAKENEWS and CLICKBAIT (the most helpful dataset) to RUMOR. Examples from FAKENEWS are long documents with a mix of formal and sensational styles, whereas CLICKBAIT contains short, sensational sentences.

However, as the model is trained on more datasets, adding the less similar FAKENEWS task actually *improves* overall performance (90.5 \rightarrow 92.5 F1 in three datasets), despite hurting the model trained on RUMOR only (86.9 \rightarrow 78.7 F1). We hypothesize this is due, in part, to including more diverse sources of data, which improves the *robustness* of the model to different types of misinformation.

4 Generalizability Analysis

New types, domains, and subjects of misinformation arise frequently. Promptly responding to these new sources is challenging, as they can spread widely before there is time to collect sufficient task-specific training examples. For instance, the rapid spread of COVID-19 was accompanied by equally fast spread of large quantities of misinformation (Joszt, 2020; Kouzy et al., 2020).

³Other datasets show similar findings.

Therefore, we carry out experiments to evaluate the generalization ability of UNIFIEDM2 representation to unseen misinformation (i) tasks/datasets and (ii) events. The first experiment is about fast adaption ability (few-shot training) to handle a new task/dataset, whereas the second experiment is about the model’s ability to perform well on events unseen during training.

4.1 Unseen Task/Dataset Generalizability

Dataset We evaluate using the following four unseen datasets: PROPAGANDA (Da San Martino et al., 2019), which contains 21,230 propaganda and non-propaganda sentences, with the propaganda sentences annotated by fine-grained propaganda technique labels, such as “Name calling” and “Appeal to fear”; POLITIFACT (Shu et al., 2019), which contains 91 true and 91 fake news *articles* collected from PolitiFact’s fact-checking platform; BUZZFEED (Shu et al., 2019), which contains 120 true and 120 fake news *headlines* collected from BuzzFeed’s fact-checking platform; and COVIDTWITTER (Alam et al., 2020), which contains 504 COVID-19-related tweets. For our experiment, we use two of the annotations: 1) *Twitter Check-worthiness*: does the tweet contain a verifiable factual claim? 2) *Twitter False Claim*: does the tweet contain false information?

Few-shot Experiments We compare the few-shot performance of UNIFIEDM2 against off-the-shelf RoBERTa and single-task RoBERTa. For each unseen dataset, a new MLP classification head is trained on top of the RoBERTa encoder, in a few-shot manner. Given N_d to be the size of the given dataset d , we train the few-shot classifiers with k randomly selected samples and evaluate on the remaining $N - k$ samples. We test with $k = 10, 25, 50$. Note that for single-task RoBERTa, we report the average performance across the four

Model	Acc	F1
SoTA'19 (Li et al., 2019)	48.30%	41.80%
SoTA'20 (Yu et al., 2020)	39.60%	46.60%
Vanilla	47.07%	33.90%
UNIFIEDM2	64.74%	44.74%

Table 5: Average acc and macro-F1 scores from leave-one-event-out cross-validation setup for RUMOR task.

task-specific models (ST average).

As shown in Table 4, our UNIFIEDM2 encoder can quickly adapt to new tasks, even with very little in-domain data. While both the single-task models and UNIFIEDM2 significantly outperform vanilla RoBERTa, UNIFIEDM2 further outperforms the single-task models, indicating that multi-task learning can aid task generalizability.

4.2 Unseen Event Generalizability

Dataset We use the previously introduced RUMOR dataset, which includes nine separate events, for this experiment. A group of works (Kochkina et al., 2018; Li et al., 2019; Yu et al., 2020) have used this dataset in a leave-one-event-out cross-validation setup (eight events for training and one event for testing) to take event generalizability into consideration in their model evaluation. We conduct a supplementary experiment following this evaluation setup for the completeness of our analysis.

Experiment First, we train the UNIFIEDM2 encoder *without* RUMOR data, and then fine-tune and evaluate in the leave-one-event-out cross-validation setup. Note that we re-train the UNIFIEDM2 encoder to ensure that it has no knowledge of the left-out-event testset. Results in Table 5 show that our proposed method outperforms two recent SoTA models (Li et al., 2019; Yu et al., 2020) by an absolute 16.44% and 25.14% in accuracy. This indicates that unified misinformation representations are helpful in event generalizability as well.

5 Related Work

Existing misinformation works take three main approaches: *Content-based* approaches examine the language of a document only. Prior works have looked at linguistic features such as hedging words and emotional words (Rubin et al., 2016; Potthast et al., 2018; Rashkin et al., 2017; Wang, 2017). *Fact-based* approaches leverage evidence from external sources (e.g., Wikipedia, Web) to determine

the truthfulness of the information (Etzioni et al., 2008; Wu et al., 2014; Ciampaglia et al., 2015; Popat et al., 2018; Thorne et al., 2018; Nie et al., 2019). Finally, *social-data-based* approaches use the surrounding social data—such as the credibility of the authors of the information (Long et al., 2017; Kirilin and Strube, 2018; Li et al., 2019) or social engagement data (Derczynski et al., 2017; Ma et al., 2018; Kwon et al., 2013; Volkova et al., 2017).

Though prior works have explored multi-task learning within misinformation, they have focused exclusively on one domain. These works try to predict two different labels on the same set of examples from a single (Kochkina et al., 2018) or two closely-related datasets (Wu et al., 2019). In contrast, our proposed approach crosses not just *task* or *dataset* boundaries, but also *format* and *domain* boundaries. Furthermore, prior works focus on using an auxiliary task to boost the performance of the main task, while we focus on using multitasking to *generalize* across many domains. Thus, the focus of this work is not the multitask paradigm, but rather the unification of the various domains, using multitasking.

6 Conclusion

In this paper, we introduced UNIFIEDM2, which unifies multiple domains of misinformation with a single multi-task learning setup. We empirically showed that such unification improves the model’s performance against strong baselines, and achieves new state-of-the-art results. Furthermore, we show that UNIFIEDM2 can generalize to out-of-domain misinformation tasks and events, and thus can serve as a good starting point for others working on misinformation.

References

- Firoj Alam, Fahim Dalvi, Shaden Shaar, Nadir Durrani, Hamdy Mubarak, Alex Nikolov, Giovanni Da San Martino, Ahmed Abdelali, Hassan Sajjad, Kareem Darwish, et al. 2020. Fighting the COVID-19 infodemic in social media: A holistic perspective and a call to arms. *arXiv preprint arXiv:2007.07996*.
- Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.
- Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. *PloS one*, 10(6):e0128193.

- Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeno, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5640–5650.
- Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 69–76.
- Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. [Open information extraction from the Web](#). *Commun. ACM*, 51(12):68–74.
- Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. [In plain sight: Media bias through the lens of factual reporting](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6344–6350.
- Laura Joszt. 2020. [Combatting covid-19 misinformation](#).
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Angelika Kirilin and Micheal Strube. 2018. Exploiting a speaker’s credibility to detect fake news. In *Proceedings of Data Science, Journalism & Media workshop at KDD (DSJM’18)*.
- Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3402–3413.
- Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, and Khalil Baddour. 2020. Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, 12(3).
- Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108. IEEE.
- Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor detection by exploiting user credibility information, attention and multi-task learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1173–1179.
- Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019a. [Multi-task deep neural networks for natural language understanding](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496, Florence, Italy. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. Fake news detection through multi-perspective speaker profiles. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 252–256.
- Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1980–1989.
- Bernard Marr. 2018. [How much data do we create every day? the mind-blowing stats everyone should read](#). *Forbes*.
- Andreas Maurer, Massimiliano Pontil, and Bernardino Romera-Paredes. 2016. The benefit of multitask representation learning. *The Journal of Machine Learning Research*, 17(1):2853–2884.
- Yixin Nie, Haonan Chen, and Mohit Bansal. 2019. Combining fact extraction and verification with neural semantic matching networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6859–6866.
- Amin Omidvar, Hui Jiang, and Aijun An. 2018. Using neural network for identifying clickbaits in online news media. In *Annual International Symposium on Information Management and Big Data*, pages 220–232. Springer.
- Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. Declare: Debunking fake news and false claims using evidence-aware deep learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 22–32.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. [A stylistic inquiry into hyperpartisan and fake news](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 231–240, Melbourne, Australia. Association for Computational Linguistics.

- Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937.
- Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pages 7–17, San Diego, California. Association for Computational Linguistics.
- Sebastian Ruder. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.
- Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320.
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 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)*, pages 809–819.
- Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 647–653.
- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.
- William Yang Wang. 2017. “Liar, Liar Pants on Fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.
- Lianwei Wu, Yuan Rao, Haolin Jin, Ambreen Nazir, and Ling Sun. 2019. Different absorption from the same sharing: Sifted multi-task learning for fake news detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4636–4645.
- You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2014. Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7):589–600.
- Jianfei Yu, Jing Jiang, Ling Min Serena Khoo, Hai Leong Chieu, and Rui Xia. 2020. Coupled hierarchical transformer for stance-aware rumor verification in social media conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1392–1401, Online. Association for Computational Linguistics.
- Yu Zhang and Qiang Yang. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*.
- Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989.