QCRI at SemEval-2023 Task 3: News Genre, Framing and Persuasion Techniques Detection using Multilingual Models

Maram Hasanain[1], **Ahmed Oumar El-Shangiti**[1]*, **Rabindra Nath Nandi**[2], **Preslav Nakov**[3] and **Firoj Alam**[1]

[1]Qatar Computing Research Institute, HBKU, Qatar
[2]Hishab Singapore Pte. Ltd, Singapore, [3]MBZUAI, UAE,
maramhasanain@gmail.com, ahmedmohamedlemin@gmail.com,
rabindro.rath@gmail.com, preslav.nakov@mbzuai.ac.ae, fialam@hbku.edu.qa

Abstract

Misinformation spreading in mainstream and social media has been misleading users in different ways. Manual detection and verification efforts by journalists and fact-checkers can no longer cope with the great scale and quick spread of misleading information. This motivated research and industry efforts to develop systems for analyzing and verifying news spreading online. The SemEval-2023 Task 3 is an attempt to address several subtasks under this overarching problem, targeting writing techniques used in news articles to affect readers' opinions. The task addressed three subtasks with six languages, in addition to three "surprise" test languages, resulting in 27 different test setups. This paper describes our participating system to this task. Our team is one of the 6 teams that successfully submitted runs for all setups. The official results show that our system is ranked among the top 3 systems for 10 out of the 27 setups.

1 Introduction

Monitoring and analyzing news have become an important process to understand how different topics (e.g., political) are reported in different news media and within and across countries. This has many important applications since the tone, framing, and factuality of news reporting can significantly affect public reactions toward social or political agendas. A news piece can be manipulated on multiple aspects to sway readers' perceptions and actions. Going beyond information factuality, other aspects include objectivity/genre, framing dimensions inserted to steer the focus of the audience (Card et al., 2015), and propaganda techniques used to persuade readers towards a certain agenda (Barrón-Cedeno et al., 2019; Da San Martino et al., 2019a).

News categorization is a well studied problem in the natural language processing field. Recently, research attention has focused on classifying news by factuality (Zhou and Zafarani, 2020; Nakov et al., 2021), or other related categorizations such as fake vs. satire news (Low et al., 2022; Golbeck et al., 2018). However, there have been efforts towards other classification dimensions. Card et al. (2015) developed a corpus of news articles annotated by 15 framing dimensions such as economy, capacity and resources, and fairness and equality, to support development of systems for news framing classification. Moreover, identifying propagandistic content has gained a lot of attention over several domains including news (Barrón-Cedeno et al., 2019; Da San Martino et al., 2019a), social media (Alam et al., 2022) and multimodal content (Dimitrov et al., 2021a,b).

The SemEval-2023 Task 3 shared task aims at motivating research in the aforementioned categorization tasks, namely: detection and classification of the *genre*, *framing*, and the *persuasion techniques* in news articles (Piskorski et al., 2023). It targets multiple languages including English, French, German, Italian, Polish, and Russian to push the research on multilingual systems. Moreover, to promote development of language-agnostic models, the task organizers released test subsets for three surprise languages (Georgian, Greek, and Spanish).

Our proposed system is based on fine-tuning transformer based models (Vaswani et al., 2017) in multiclass and multi-label classification settings for different tasks and languages. We participated in all three subtasks submitting runs for all nine languages, which resulted in 27 testing setups. We experimented with different mono and multilingual transformer models, such as BERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020; Chi et al., 2022) among others. In addition, we also experimented with data augmentation.

The rest of the paper is organized as follows. Section 2 gives an overview of related work. In section 3, we present the proposed system. In section 4, we provide the details of our experiments.

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 1237–1244
July 13-14, 2023 ©2023 Association for Computational Linguistics

Section 5 presents the results for our official runs, and finally, we conclude our paper in section 6.

2 Related Work

2.1 News Genre Categorization

Prior works on automated news categorization have focused on various aspects such as topic, style, how news is presented or structured, and intended audience (Einea et al., 2019; Chen and Choi, 2008; Yoshioka et al., 2001; Stamatatos et al., 2000). News articles have also been categorized based on their factuality and deceptive intentions (Golbeck et al., 2018). For example, fake news is false and the intention is deceive where satire news is also false but the intent is not deceive rather to call out, ridicule, or expose behavior that is shameful, corrupt, or otherwise "bad".

2.2 Propaganda Detection

Propaganda is defined as the use of automatic approaches to intentionally disseminate misleading information over social media platforms (Woolley and Howard, 2018). Recent work on propaganda detection has focused on news articles (Barrón-Cedeno et al., 2019; Rashkin et al., 2017; Da San Martino et al., 2019b, 2020), multimodal content such as memes (Dimitrov et al., 2021a,b) and tweets (Vijayaraghavan and Vosoughi, 2022; Alam et al., 2022). Several annotated datasets have been developed for the task such as `TSHP-17` (Rashkin et al., 2017), and `QProp` (Barrón-Cedeno et al., 2019). Habernal et al. (2017, 2018) developed a corpus with 1.3k arguments annotated with five fallacies (e.g., red herring fallacy), which directly relate to propaganda techniques. Da San Martino et al. (2019b) developed a more fine-grained taxonomy consisting of 18 propaganda techniques with annotation of news articles. Moreover, the authors proposed a multigranular deep neural network that captures signals from the sentence-level task and helps to improve the fragment-level classifier. An extended version of the annotation scheme was proposed to capture information in multimodal content (Dimitrov et al., 2021a). Datasets in languages other than English have been proposed. For example, using the same annotation scheme from (Dimitrov et al., 2021a), Alam et al. (2022) developed a dataset of Arabic tweets and organized a shared task on Arabic propaganda technique detection. Vijayaraghavan and Vosoughi (2022) developed a dataset of tweets, which are weakly labeled with different fine-grained propaganda techniques. They also proposed a neural approach for classification.

2.3 Framing

Framing refers to representing different salient aspects and perspectives for the purpose of conveying the latent meaning about an issue (Entman, 1993). Recent work on automatically identifying media frames includes developing coding schemes and semi-automated methods (Boydstun et al., 2013), datasets such as the Media Frames Corpus (Card et al., 2015), systems to automatically detect media frames (Liu et al., 2019a; Zhang et al., 2019), large-scale automatic analysis of news articles (Kwak et al., 2020), and semi-supervised approaches (Cheeks et al., 2020).

Given the multilingual nature of the datasets released with the task at hand, our work is focused on designing a multilingual approach for news classification for the three subtasks of interest.

3 System Overview

Our system is comprised of preprocessing followed by fine-tuning pre-trained transformer models. The preprocessing part includes standard model specific tokenization. Our experimental setup consists of *(i)* monolingual ($*_{\text{mono}}$): training and evaluating monolingual transformer model for each language and subtask; *(ii)* multilingual ($*_{\text{multi}}$): combining subtask specific data from all languages for training, and evaluating the model on task and language specific data; *(iii)* data augmentation ($*_{\text{aug}}$): applying data augmentation on language specific training set, then training a monolingual model using augmented dataset, and evaluating it on the test set. This has been applied for each subtask.

3.1 Data Augmentation

Data augmentation is an effective way to deal with class imbalance issues or to increase the size of the training dataset or increase within-class variation. Typically, textual data augmentation has been done by upsampling techniques such as SMOTE (Chawla et al., 2002), however, that approach is applied to the vector representation. Very recently, some useful strategies are introduced for textual data augmentation (Feng et al., 2021), which range from rule-based approaches to model-based techniques. Wei and Zou (2019) proposed a set of token-level random perturbation operations including random insertion, deletion, and swap, which have been employed in several studies (Feng et al., 2021; Alam et al., 2020).

We used such approaches with contextual representation from transformer models in this study. These include (i) synonym augmentation using WordNet, (ii) word insertion and substitution using BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019b) and DistilBERT (Sanh et al., 2019). More details on the implementation of these approaches can be found in the following data augmentation package.[1]

4 Experiments

In this section, we describe the tasks and datasets used during experiments and provide implementation details for our models.

4.1 Task and Dataset

The SemEval-2023 Task 3 is composed of 3 subtasks for each language:

1. **News Genre Categorization (*subtask1*)**: Given a news article in a particular language, classify it to an *opinion*, *news reporting*, or a *satire* piece. This is a multiclass classification task at the article level.

2. **Framing Detection (*subtask2*)**: Given a news article, identify the frames used in the article. This is a multi-label classification task at the article level. This task includes 14 frames/labels such as *economic*, *capacity and resources*, *morality*, and *fairness and equality*.

3. **Persuasion Techniques Detection (*subtask3*)**: Given an article, identify the persuasion technique(s) present in each paragraph. This is a multi-label classification task at the paragraph level. This task includes 23 techniques/labels such as *loaded language*, *appeal to authority*, *appeal to popularity*, and *appeal to values.*

The task organizers released three subsets (train, development and test) of data per language of the six main languages for each subtask. Further details and statistics can be found in (Piskorski et al., 2023). Starting with the six *train* subsets, we apply three methods to acquire new versions of these train subsets:

[1] `https://github.com/makcedward/nlpaug`

HF Model Name	**Language**
xlm-roberta-large	Multilingual
bert-large-cased	English
roberta-large	English
dbmdz/bert-base-french-europeana-cased	French
dbmdz/bert-base-german-uncased	German
uklfr/gottbert-base	German
dbmdz/bert-base-italian-uncased	Italian
sdadas/polish-roberta-large-v2	Polish
allegro/herbert-large-cased	Polish
DeepPavlov/rubert-base-cased	Russian

Table 1: Pre-trained models used in experiments. For languages with multiple models, the best ones are shown in bold, which are also comparable in the monolingual training setup on the dev subset across all three subtasks.

1. Train subset splitting: we randomly split each of the train subsets into 80-20 splits to acquire training and validation subsets for each subtask and each language. As will be shown in the following subsection, our models were re-trained using different random seeds. The validation set is used to select the random seed leading to the best model.

2. Multilingual dataset construction: to support our multilingual training setup, we combine the training subsets resulting from the previous step for all languages to create a multilingual training subset. We apply the same approach to the validation subsets.

3. Data augmentation: for each of our generated training splits, we apply data augmentation to it and use the resulting datasets to train a monolingual model for each subtask and each language.

4.2 Implementation Details

We use HuggingFace (HF) library (Wolf et al., 2020) on top of PyTorch framework (Paszke et al., 2017) as our base and source of all the pre-trained language models. Since different random initialization can considerably affect the model performance, we train the model for each language with k different random seeds.

For all experiments, we use Adam optimizer (Kingma and Ba, 2015) with the learning rate of $2{\times}10^{-5}$. In setting other parameters of the models, we distinguish between *subtask1* and *subtask2* that operate on the document level, and *subtask3*$_{\text{multi/aug}}$ that works at the paragraph level and has a much larger training subset. Only for

Lang	Rank	Run	$\mathbf{F1_{macro}}$	$\mathbf{F1_{micro}}$
EN	1	MELODI	0.784	0.815
	16	Baseline	0.288	0.611
	17	$\text{QCRI}_{\text{multi}}$	0.281	0.593
FR	1	UMUTeam	0.835	0.880
	2	QCRI_{aug}	0.767	0.800
	10	Baseline	0.568	0.740
GE	1	UMUTeam	0.820	0.820
	1	SheffieldVeraAI	0.820	0.820
	7	$\text{QCRI}_{\text{mono}}$	0.667	0.660
	9	Baseline	0.630	0.760
IT	1	Hitachi	0.768	0.852
	7	$\text{QCRI}_{\text{mono}}$	0.541	0.787
	12	Baseline	0.389	0.672
PO	1	FTD	0.786	0.936
	10	$\text{QCRI}_{\text{mono}}$	0.571	0.830
	13	Baseline	0.490	0.830
RU	1	Hitachi	0.755	0.750
	6	$\text{QCRI}_{\text{multi}}$	0.567	0.653
	12	Baseline	0.398	0.653
KA	1	Riga	1.000	1.000
	4	$\text{QCRI}_{\text{multi}}$	0.622	0.897
	13	Baseline	0.256	0.345
GR	1	SinaaAI	0.806	0.813
	4	$\text{QCRI}_{\text{multi}}$	0.708	0.813
	15	Baseline	0.171	0.344
ES	1	DSHacker	0.563	0.567
	3	$\text{QCRI}_{\text{multi}}$	0.489	0.567
	16	Baseline	0.154	0.300

Table 2: Official results for all nine test languages in *subtask1*. $\mathbf{F1_{macro}}$ is the official evaluation measure for this subtask. Subscripts for our team runs indicate the training setup used.

Lang	Rank	Run	$\mathbf{F1_{micro}}$	$\mathbf{F1_{macro}}$
EN	1	SheffieldVeraAI	0.579	0.539
	7	$\text{QCRI}_{\text{multi}}$	0.513	0.419
	18	Baseline	0.350	0.274
FR	1	MarsEclipse	0.553	0.537
	7	$\text{QCRI}_{\text{multi}}$	0.480	0.430
	15	Baseline	0.329	0.276
GE	1	MarsEclipse	0.711	0.660
	2	$\text{QCRI}_{\text{multi}}$	0.660	0.606
	17	Baseline	0.487	0.418
IT	1	MarsEclipse	0.617	0.545
	2	$\text{QCRI}_{\text{multi}}$	0.599	0.479
	13	Baseline	0.486	0.372
PO	1	MarsEclipse	0.673	0.638
	3	$\text{QCRI}_{\text{multi}}$	0.642	0.599
	10	Baseline	0.594	0.532
RU	1	MarsEclipse	0.450	0.303
	3	$\text{QCRI}_{\text{multi}}$	0.434	0.364
	13	Baseline	0.230	0.218
KA	1	SheffieldVeraAI	0.654	0.679
	6	$\text{QCRI}_{\text{multi}}$	0.517	0.457
	13	Baseline	0.260	0.251
GR	1	SheffieldVeraAI	0.546	0.454
	6	$\text{QCRI}_{\text{multi}}$	0.519	0.400
	13	Baseline	0.345	0.057
ES	1	mCPT	0.571	0.455
	6	$\text{QCRI}_{\text{multi}}$	0.488	0.390
	17	Baseline	0.120	0.095

Table 3: Official results for all nine test languages in *subtask2*. $\mathbf{F1_{micro}}$ is the official evaluation measure for this subtask. Subscripts for our team runs indicate the training setup used.

subtask3$_{\text{multi/aug}}$, the number of epochs=5, k=5, maximum sequence length=256, and batch size=8. For all remaining training setups and subtasks, the number of epochs=10, k=10, maximum sequence length=512, and batch size=4.

For each of the three training setups described in section 3, the models trained using k seeds for a language are evaluated over our validation subset using the official evaluation measure for the corresponding subtask. The model with the best performance is then applied to the development set. Eventually, the training setup that has the best performance on the development subset will be used to generate the official run for the corresponding subtask and test language. As for the “surprise” test languages, we use the model trained on the multilingual training subset with the best performance on the multilingual validation subset.

For our multilingual training setup, we opt to use XLM-RoBERTa (Conneau et al., 2020). As for all other setups, we used per-language monolingual pre-trained models listed in Table 1.

5 Results

The results for our official runs per subtask are shown in Tables 2, 3 and 4. For each subtask, we compare our official runs to two baselines: the top run in each test language, and the baseline as reported by the task organizers.

We observe that the multilingual models are generally the best performing models across all tasks. On average, the performance of the system was best for *subtask3* with a slight average ranking difference compared to *subtask2*. Another interesting observation is that although *subtask3* has much larger train subsets , since it operates on the paragraph level, this did not improve the average system ranking across languages when compared to *subtask2*. The results also clearly show the robustness of our model across languages and subtasks, as it

Lang	Rank	Run	$F1_{micro}$	$F1_{macro}$
EN	1	APatt	0.376	0.129
	8	$QCRI_{multi}$	0.320	0.133
	19	Baseline	0.195	0.069
FR	1	NAP	0.469	0.322
	5	$QCRI_{multi}$	0.401	0.226
	16	Baseline	0.240	0.099
GE	1	KInITVeraAI	0.513	0.233
	3	$QCRI_{multi}$	0.498	0.231
	17	Baseline	0.317	0.083
IT	1	KInITVeraAI	0.550	0.214
	6	$QCRI_{multi}$	0.513	0.209
	16	Baseline	0.397	0.122
PO	1	KInITVeraAI	0.430	0.179
	5	$QCRI_{multi}$	0.378	0.156
	18	Baseline	0.179	0.059
RU	1	KInITVeraAI	0.387	0.189
	3	$QCRI_{multi}$	0.361	0.182
	15	Baseline	0.207	0.086
KA	1	KInITVeraAI	0.457	0.328
	2	$QCRI_{multi}$	0.414	0.339
	14	Baseline	0.138	0.141
GR	1	KInITVeraAI	0.267	0.126
	2	$QCRI_{multi}$	0.265	0.129
	14	Baseline	0.088	0.006
ES	1	TeamAmpa	0.381	0.244
	4	$QCRI_{multi}$	0.350	0.157
	11	Baseline	0.248	0.020

Table 4: Official results for all nine test languages in *subtask3*. $\mathbf{F1_{micro}}$ is the official evaluation measure for this subtask. Subscripts for our team runs indicate the training setup used.

managed to be among the best 3 runs for 10 out of the 27 test subsets, and it was among the top 5 runs for 15 of them.

Results over *subtask1* and *subtask3* showed that our proposed system had a strong cross-lingual transfer ability when training the model on multilingual data and testing it on unseen languages (Georgian, Greek and Spanish).

6 Conclusion

In this paper, we presented our experiments and findings on news genre categorization, framing and persuasion techniques detection on multiple languages, which was a part of SemEval-2023 Task 3 shared task. The task includes 27 test setups for three subtasks and nine test languages. Our team successfully submitted runs for all setups. We proposed a system that is based on fine-tuning transformer models in multiclass and multi-label classification settings. We experimented with different mono and multilingual pre-trained models, in addition to data augmentation. From the experimental results, we observed that our multilingual model based on XLM-RoBERTa performs better across all tasks, even on unseen languages.

Our future work includes domain adaptation and further exploration of data augmentation techniques.

Acknowledgments

This publication was made possible by NPRP grant 14C-0916-210015 *The Future of Digital Citizenship in Qatar: a Socio-Technical Approach* from the Qatar National Research Fund.

Part of this work was also funded by Qatar Foundation's IDKT Fund TDF 03-1209-210013: *Tanbih: Get to Know What You Are Reading*.

The views, opinions, and findings presented in this paper are those of the authors alone and do not necessarily reflect the views, policies, or positions of the QNRF or any other affiliated organizations.

References

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Tanvirul Alam, Akib Khan, and Firoj Alam. 2020. Punctuation restoration using transformer models for high- and low-resource languages. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 132–142.

Alberto Barrón-Cedeno, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5):1849–1864.

Amber E Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *New Directions in Analyzing Text as Data Workshop*, London School of Economics and Political Science, London, UK.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th*

International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 438–444, Beijing, China. Association for Computational Linguistics.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Loretta H Cheeks, Tracy L Stepien, Dara M Wald, and Ashraf Gaffar. 2020. Discovering news frames: An approach for exploring text, content, and concepts in online news sources. In *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications*, pages 702–721. IGI Global.

Guangyu Chen and Ben Choi. 2008. Web page genre classification. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 2353–2357.

Zewen Chi, Shaohan Huang, Li Dong, Shuming Ma, Bo Zheng, Saksham Singhal, Payal Bajaj, Xia Song, Xian-Ling Mao, Heyan Huang, and Furu Wei. 2022. XLM-E: Cross-lingual language model pre-training via ELECTRA. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6170–6182, Dublin, Ireland. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 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.

Giovanni Da San Martino, Alberto Barrón-Cedeño, and Preslav Nakov. 2019a. Findings of the NLP4IF-2019 shared task on fine-grained propaganda detection. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 162–170, Hong Kong, China. Association for Computational Linguistics.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI '20, pages 4826–4832.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019b. 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 '19, pages 5636–5646, Hong Kong, China.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT '19, pages 4171–4186, Minneapolis, Minnesota, USA.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021a. Detecting propaganda techniques in memes. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, ACL-IJCNLP '21, pages 6603–6617.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021b. SemEval-2021 task 6: Detection of persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation*, SemEval '21, pages 70–98.

Omar Einea, Ashraf Elnagar, and Ridhwan Al Debsi. 2019. Sanad: Single-label arabic news articles dataset for automatic text categorization. *Data in Brief*, 25:104076.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

Steven Y Feng, Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. 2021. A survey of data augmentation approaches for NLP. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 968–988, Online. Association for Computational Linguistics.

Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jennine B Everett, et al. 2018. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science*, pages 17–21.

Ivan Habernal, Raffael Hannemann, Christian Pollak, Christopher Klamm, Patrick Pauli, and Iryna Gurevych. 2017. Argotario: Computational argumentation meets serious games. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 7–12, Copenhagen, Denmark.

Ivan Habernal, Patrick Pauli, and Iryna Gurevych. 2018. Adapting serious game for fallacious argumentation to German: Pitfalls, insights, and best practices. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, LREC '18, pages 3329–3335, Miyazaki, Japan.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million New York Times articles from 2000 to 2017. In *Proceedings of the 12th ACM Conference on Web Science*, WebSci '20, pages 305–314, Southampton, United Kingdom.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019a. Detecting frames in news headlines and its application to analyzing news framing trends surrounding US gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning*, CoNLL '19, pages 504–514, Hong Kong, China.

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:1907.11692*.

Jwen Fai Low, Benjamin C.M. Fung, Farkhund Iqbal, and Shih-Chia Huang. 2022. Distinguishing between fake news and satire with transformers. *Expert Systems with Applications*, 187:115824.

Preslav Nakov, Husrev Taha Sencar, Jisun An, and Haewoon Kwak. 2021. A survey on predicting the factuality and the bias of news media. *arXiv preprint arXiv:2103.12506*.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *NIPS-W*.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation*, SemEval 2023, Toronto, Canada.

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*, EMNLP '17, pages 2931–2937, Copenhagen, Denmark.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Efstathios Stamatatos, Nikos Fakotakis, and George Kokkinakis. 2000. Automatic text categorization in terms of genre and author. *Computational linguistics*, 26(4):471–495.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Prashanth Vijayaraghavan and Soroush Vosoughi. 2022. TWEETSPIN: Fine-grained propaganda detection in social media using multi-view representations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3433–3448, Seattle, United States. Association for Computational Linguistics.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. 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 6382–6388.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 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.

Samuel C Woolley and Philip N Howard. 2018. *Computational propaganda: political parties, politicians, and political manipulation on social media*. Oxford University Press.

Takeshi Yoshioka, George Herman, JoAnne Yates, and Wanda Orlikowski. 2001. Genre taxonomy: A knowledge repository of communicative actions. *ACM transactions on information systems (TOIS)*, 19(4):431–456.

Yifan Zhang, Giovanni Da San Martino, Alberto Barrón-Cedeño, Salvatore Romeo, Jisun An, Haewoon Kwak, Todor Staykovski, Israa Jaradat, Georgi Karadzhov, Ramy Baly, Kareem Darwish, James Glass, and Preslav Nakov. 2019. Tanbih: Get to know what you are reading. 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): System Demonstrations*, pages 223–228, Hong Kong, China. Association for Computational Linguistics.

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