Frank at ArAIEval Shared Task: Arabic Persuasion and Disinformation: The Power of Pretrained Models

Dilshod Azizov[1], Jiyong Li[2], Shangsong Liang[1,*]
[1]Mohamed bin Zayed University of Artificial Intelligence
[2]Sun Yat-sen University
dilshod.azizov@mbzuai.ac.ae, lijy373@mail2.sysu.edu.cn
[*]Corresponding author, liangshangsong@gmail.com

Abstract

In this work, we present our systems developed for "ARAIEVAL" shared task of ArabicNLP 2023 (Hasanain et al., 2023a). We used an mBERT transformer for Subtask 1A, which targets persuasion in Arabic tweets, and we used the MARBERT transformer for Subtask 2A to identify disinformation in Arabic tweets. Our persuasion detection system achieved micro-F1 of **0.745** by surpassing the baseline by 13.2%, and registered a macro-F1 of 0.717 based on leaderboard scores. Similarly, our disinformation system recorded a micro-F1 of **0.816**, besting the naïve majority by 6.7%, with a macro-F1 of 0.637. Furthermore, we present our preliminary results on a variety of pre-trained models. In terms of overall ranking, our systems placed 7th out of 16 and 12th out of 17 teams for Subtasks 1A and 2A, respectively.

1 Introduction

The digital communication landscape, vast in its dynamism, constantly evolves, presenting unique challenges in diverse cultural and linguistic contexts. Arabic, with its rich historical and poetic traditions, spoken by more than 420 million people, is no exception (Qu et al., 2023). In the age of digital connectivity, platforms like Twitter have become a boon and a bane. They enable rapid dissemination of information, but also disinformation spread, which can manipulate public perceptions and cause socio-political instability (Raj and Goswami, 2020). Because tweets are so brief, accuracy is essential. This requires the use of strong rhetorical elements, making them an ideal environment for disinformation and persuasive strategies (Hasanain et al., 2023b; Hardalov et al., 2021; Nakov and Da San Martino, 2021).

Taking into account the vast diversity of Arabic dialects and cultural nuances, identifying these strategies is challenging, especially with the rise in misinformation campaigns (Dimitrov et al., 2021). The importance of addressing this issue is amplified by the geopolitical significance of Arabic-speaking regions, where digital narratives can influence diplomacy, policy decisions, and public sentiment (Al-Rawi et al., 2022; Guellil et al., 2021; Cheng et al., 2021).

In response, the ARAIEVAL shared task of ArabicNLP 2023 focuses on two critical areas:

Subtask 1A: *Given a multigenre (tweet and news paragraphs of the news articles) snippet, identify whether it contains content with persuasion technique.*

Subtask 2A: *Given a tweet, categorize whether it is disinformative (Hasanain et al., 2023a).*

Our paper offers the following contributions:

- We propose systems that use mBERT (Devlin et al., 2018) for persusaion detection and MARBERT (Abdul-Mageed et al., 2020) for disinfromation identification.
- We compare the performance of mBERT *vs.* XLM-RoBERTa (Liu et al., 2019) and MARBERT for subtask 1A. In Subtask 2A, we compare MARBERT *vs.* AraBERT(Antoun et al., 2020), and ALBERT (Lan et al., 2019).

In Section 2, we outline previous and more recent studies on the identification of persuasion and disinformation. In Section 3, we illustrate a thorough examination of the dataset. In Section 4 we describe the systems and the results. Finally, Section 5 presents our conclusion and suggests directions for future exploration.

2 Related Work

In recent years, Natural Language Processing (NLP) has experienced significant advances, particularly in detecting persuasive techniques and misinformation across various languages. Historically, English-centric models have been at the forefront, showcasing breakthroughs in understanding and

Proceedings of the The First Arabic Natural Language Processing Conference (ArabicNLP 2023), pages 583–588
December 7, 2023 ©2023 Association for Computational Linguistics

	Train	Dev	Test	Total
Subtask 1A	2,120	566	503	3,189
Subtask 2A	14,147	2,115	3,739	20,002

Table 1: Statistics about distributon of tweets in train/dev/test of Subtasks 1A and 2A

auto-detecting persuasion (Haouari et al., 2020; Piskorski et al., 2023). On the contrary, Arabic, characterized by its linguistic diversity and cultural richness, has seen relatively limited focus in the domain of persuasion detection. While the foundational research of Arabic NLP revolved around sentiment analysis (Abdulla et al., 2013) and stance detection (Almiman et al., 2020; Hardalov et al., 2021), the nuanced domain of detecting persuasion techniques in Arabic remained underexplored due to the complex morphology of the language and the diverse dialects (Alam et al., 2022).

Furthermore, recent studies, such as Al-Sallab et al. (2017), have emphasized the need for advanced embeddings and specialized datasets tailored to Arabic peculiarities. An emerging trend marries traditional Arabic linguistic studies with contemporary machine learning, targeting the precise detection of persuasive techniques in Arabic content (Alam et al., 2022).

The rapid proliferation of disinformation in the digital age, especially via social media platforms, requires the inter-related studies on persuasion and misinformation studies (Peng et al., 2023). In particular, CheckThat! lab at CLEF has embarked on multifaceted research on misinformation in different languages, encompassing fact-checking, checkworthiness, bias identification, and source credibility assessment (Da San Martino et al., 2023; Azizov et al., 2023; Nakov et al., 2023; Barrón-Cedeño et al., 2023; Barrón-Cedeño et al., 2023; Elsayed et al., 2019a,b; Hasanain et al.; Barrón-Cedeño et al.; Nakov et al., 2021a,b). This is consistent with the contemporary research trends that have changed from analyzing only news articles to scrutinizing social media for propaganda detection (Woolley and Howard, 2018; Martino et al., 2020b). Interestingly, another study by (Zhang et al., 2019) proposes a Bayesian deep learning model for misinformation detection, incorporating claim responses and quantifying prediction uncertainty, achieving superior performance in public datasets.

Moreover, Da San Martino et al. (2019) delved deeply into persuasive techniques, highlighting the emotional signals that resonate with readers. This foundational work paved the way for subsequent endeavors, notably the "Detection of Propaganda Techniques in News Articles" challenge posited by (Martino et al., 2020a). Building on this momentum, a recent investigation by (Mubarak et al., 2023) sought to discern and categorize the underlying reasons for the deletion of Arabic tweets, and later designed predictive models for potential deletions. In the multimedia realm, Dimitrov et al. (2021) emphasized the importance of detecting propaganda within memes, thus underscoring the convergence of text and imagery in disinformation campaigns.

3 Data

In this section, a detailed description of the dataset released by the ARAIEVAL shared task organizers is provided. Our primary focus is on the binary classification challenge subtasks 1A and 2A persuasion and disinformation detection dataset.

Data Attributes: *Both subtasks consist of a dataset with the same structure, comprising an ID, text, and label. However, the labels differentiate between the subtasks*

- **ID**: Numerical index of the data point.
- **Tweet for Subtask 1A**: Arabic tweet potentially containing persuasion.
- **Tweet for Subtask 2A**: Arabic tweet potentially containing disinformation .
- **Label for Subtask 1A**: “True” (indicating the presence of persuasion) and “False” (indicating the absence of persuaion).
- **Label for Subtask 2A**: “Disinfo” (denoting the text as a rumor) and “No-Disinfo” (indicating the absence of disinformation).

Dataset Size:

The dataset from ARAIEVAL is detailed in Table 1. Subtask 1A consists of less than 3.2k tweets,

	Subtask 1A		Subtask 2A	
	True	False	Disinfo	No-Disinfo
Train	1,918	509	2,656	11,491
Dev	202	57	397	1,718

Table 2: Labels distribution over the train and development set in Subtasks 1A and Subtask 2A.

while Subtask 2A contains slightly more than 20k tweets. The distribution of labels within the training and development sets can be seen in Table 2. In particular, both subtasks have an imbalance distribution of the datasets.

4 System Descriptions and Results

4.1 System Descriptions

For the assessment, we used the official evaluation tools designated for the shared task. The official measure for both subtasks is micro-F1, although the macro-F1 measure is also generated by the evaluation tools. Our models training was carried out using two NVIDIA Tesla T4 GPUs, each with 16GB memory.

Subtask 1A

mBERT. We used the mBERT base architecture. Our configuration involved a batch size of 16 and a training duration of 5 epochs with a learning rate of 5e-5. Measures were logged every 500 steps. Gradient norms were reduced to a maximum value of 1.0. ADAMW optimizer was used with a weight decay of 0.01 to mitigate overfitting. Model checkpoints were saved every 500 steps and after the end of each epoch. Both the warm-up ratio and the warm-up steps were set to zero.

Subtask 2A

MARBERT. For our binary classification task, we utilized the MARBERT base architecture, which is equipped with 12 self-attention heads, has 163M parameters and an embedding dimensionality of 768. We use the Adam optimizer with a learning rate set at 5e-5. To balance computational efficiency with model convergence, we settled on a batch size of 32. Labels "no-disinfo" and "disinfo" were encoded in 0 and 1, respectively, using a *label2id* dictionary, and decoded with a *id2label* dictionary for predictions. The training was conducted over five epochs, after which the model achieved convergence without evident signs of overfitting.

Note: For both subtasks, the data set was preprocessed using the AraBERT pre-processor and tokenizer. Text inputs were standardized to a sequence length of 512 tokens through truncation and padding.

4.2 Results

In the initial stages, we experimented with the development set, as we used it as a test set, and from the train set we cut 10% out of the total tweets for the development set. All models have been trained on 3 epochs, the rest of the hyperparameters have been used as default. Below, we dive deeper into each model's performance and postulate the reasons behind their relative successes and shortcomings.

Subtask 1A. mBERT exhibits exemplary performance in this subtask, registering the highest micro-F1 score of 0.889. Its efficiency in maintaining a balance between precision and recall is evident in its scores of 0.855 and 0.887, respectively. MARBERT closely follows with a commendable micro-F1 score of 0.881, and its precision and recall stand at 0.847 and 0.877, respectively. This suggests that while mBERT slightly edges out in terms of overall performance, MARBERT remains a strong contender. XLM-RoBERTa, although competitive, falls slightly behind with a micro-F1 score of 0.876. It has a precision score of 0.780 and a recall of 0.870, indicating that it can be more conservative in its predictions compared to the other models.

Subtask 2A. MARBERT secures the top position for this subtask with a micro-F1 score of 0.866. Its precision and recall scores are 0.856 and 0.878, respectively, indicating a balanced performance. ALBERT, with a micro-F1 score of 0.846, also shows commendable results. Its precision is slightly lower than that of MARBERT at 0.842, but it manages a recall of 0.871. mBERT has a micro-F1 score of 0.840 and exhibits similar precision and recall values of 0.840 and 0.862, respectively. This demonstrates that while MARBERT is leading in this subtask, ALBERT and mBERT remain closely competitive.

	Subtask 1A				Subtask 2A				
	Micro F1	Accuracy	Precision	Recall		Micro F1	Accuracy	Precision	Recall
mBERT	**0.889**	**0.853**	**0.855**	**0.887**	MARBERT	**0.866**	**0.854**	**0.856**	**0.878**
XLM-RoBERTa	0.876	0.781	0.780	0.870	mBERT	0.840	0.841	0.840	0.862
MARBERT	0.881	0.838	0.847	0.877	ALBERT	0.846	0.843	0.842	0.871

Table 3: The experimental results of various frameworks on the development sets of Subtasks 1A and 2A.

For the tasks at hand, our described configuration yielded notable results. In Subtask 1A, our model recorded a micro-F1 of 0.745 and a macro-F1 of 0.717. Meanwhile, for Subtask 2A, the corresponding scores were 0.816 and 0.637. The significant performance of the system can be attributed to judicious choice of models and meticulous fine-tuning. Such efforts positioned us competitively in the leaderboard rankings.

4.3 Analysis

Delving into the observed differences:

MARBERT: Continues it's impressive streak across both subtasks. It's architecture demonstrates finely tuned for classification tasks, but with the close competition in Subtask 2A, it shows that it can have specific strengths for different types of data.

mBERT: While shining brightly in Subtask 1A, it faces closer competition in Subtask 2A. Its strong recall figures hint at its efficiency in capturing most positive instances.

XLM-RoBERTa: Although trailing behind mBERT and MARBERT in Subtask 1A, its competitive scores show its capabilities. The drop in recall could suggest specific challenges in capturing all positive instances.

ALBERT: In Subtask 2A, its scores are quite competitive, especially given the close figures in the recall. This suggests that, while it may have precision challenges, it is quite adept at capturing positive instances.

Finally, after a comprehensive comparison analysis, we opt to integrate mBERT for the persuasion detection system and MARBERT for the disinformation system.

5 Conclusion and Future Work

In this paper, we discussed our approaches for the subtasks 1A and 2A of the shared task ARAIEVAL 2023 on persuasion and disinformation detection in Arabic. We employed the mBERT model for Subtask 1A and the MARBERT framework for Subtask 2A, and according to the official leaderboard results, our system achieved a micro-F1 of 0.745 and a macro-F1 of 0.717 for Subtask 1A, and a micro-F1 of 0.816 and macro-F1 of 0.637. We also detailed a series of experiments and made initial comparisons of our systems with various state-of-the-art frameworks.

In future work, we plan to delve into feature engineering, potentially integrating meta-features associated with the text, such as text length, unique word count, and sentiment analysis.

References

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2020. Arbert & marbert: deep bidirectional transformers for arabic. *arXiv preprint arXiv:2101.01785*.

Nawaf A Abdulla, Nizar A Ahmed, Mohammed A Shehab, and Mahmoud Al-Ayyoub. 2013. Arabic sentiment analysis: Lexicon-based and corpus-based. In *2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*, pages 1–6. IEEE.

Ahmed Al-Rawi, Abdelrahman Fakida, Kelly Grounds, et al. 2022. Investigation of covid-19 misinformation in arabic on twitter: Content analysis. *Jmir Infodemiology*, 2(2):e37007.

Ahmad Al-Sallab, Ramy Baly, Hazem Hajj, Khaled Bashir Shaban, Wassim El-Hajj, and Gilbert Badaro. 2017. Aroma: A recursive deep learning model for opinion mining in arabic as a low resource language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(4):1–20.

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 Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ali Almiman, Nada Osman, and Marwan Torki. 2020. Deep neural network approach for arabic community question answering. *Alexandria Engineering Journal*, 59(6):4427–4434.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*.

Dilshod Azizov, S Liang, and P Nakov. 2023. Frank at checkthat! 2023: Detecting the political bias of news articles and news media. *Working Notes of CLEF*.

Alberto Barrón-Cedeño, Firoj Alam, Tommaso Caselli, Giovanni Da San Martino, Tamer Elsayed, Andrea Galassi, Fatima Haouari, Federico Ruggeri, Julia Maria Struß, Rabindra Nath Nandi, et al. 2023. The clef-2023 checkthat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority. In *European Conference on Information Retrieval*, pages 506–517. Springer.

Alberto Barrón-Cedeño, Firoj Alam, Andrea Galassi, Giovanni Da San Martino, Preslav Nakov, , Tamer Elsayed, Dilshod Azizov, Tommaso Caselli, Gullal S. Cheema, Fatima Haouari, Maram Hasanain, Mucahid Kutlu, Chengkai Li, Federico Ruggeri, Julia Maria Struß, and Wajdi Zaghouani. 2023. Overview of the CLEF–2023 CheckThat! Lab checkworthiness, subjectivity, political bias, factuality, and authority of news articles and their source. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023)*.

Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, Shaden Shaar, and Zien Sheikh Ali. Overview of CheckThat! 2020: Automatic identification and verification of claims in social media. pages 215–236.

Qiao Cheng, Juntao Liu, Xiaoye Qu, Jin Zhao, Jiaqing Liang, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Yanghua Xiao. 2021. Hacred: A large-scale relation extraction dataset toward hard cases in practical applications. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2819–2831.

Giovanni Da San Martino, Firoj Alam, Maram Hasanain, Rabindra Nath Nandi, Dilshod Azizov, and Preslav Nakov. 2023. Overview of the clef-2023 checkthat! lab task 3 on political bias of news articles and news media. *Working Notes of CLEF*.

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 5636–5646.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021. Semeval-2021 task 6: detection of persuasion techniques in texts and images. *arXiv preprint arXiv:2105.09284*.

Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, and Pepa Atanasova. 2019a. CheckThat! at CLEF 2019: Automatic identification and verification of claims. In *Advances in Information Retrieval – European Conference on IR Research*, ECIR '19, pages 309–315.

Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, and Pepa Atanasova. 2019b. Overview of the CLEF-2019 CheckThat!: Automatic identification and verification of claims. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, LNCS, pages 301–321.

Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, and Damien Nouvel. 2021. Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences*, 33(5):497–507.

Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed. 2020. Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection. *arXiv preprint arXiv:2010.08768*.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. A survey on stance detection for mis-and disinformation identification. *arXiv preprint arXiv:2103.00242*.

Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouani, Preslav Nakov, Giovanni Da San Martino, and Abdelhakim Freihat. 2023a. Araieval shared task: Persuasion techniques and disinformation detection in arabic text. In *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)*, Singapore. Association for Computational Linguistics.

Maram Hasanain, Ahmed Oumar El-Shangiti, Rabindra Nath Nandi, Preslav Nakov, and Firoj Alam. 2023b. Qcri at semeval-2023 task 3: News genre, framing and persuasion techniques detection using multilingual models. *arXiv preprint arXiv:2305.03336*.

Maram Hasanain, Fatima Haouari, Reem Suwaileh, Zien Sheikh Ali, Bayan Hamdan, Tamer Elsayed, Alberto Barrón-Cedeño, Giovanni Da San Martino, and Preslav Nakov. Overview of CheckThat! 2020 Arabic: Automatic identification and verification of claims in social media.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut.

2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

G Martino, Alberto Barrón-Cedeno, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. Semeval-2020 task 11: Detection of propaganda techniques in news articles. *arXiv preprint arXiv:2009.02696*.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020b. A survey on computational propaganda detection. *arXiv preprint arXiv:2007.08024*.

Hamdy Mubarak, Samir Abdaljalil, Azza Nassar, and Firoj Alam. 2023. Detecting and reasoning of deleted tweets before they are posted. *arXiv preprint arXiv:2305.04927*.

Preslav Nakov, Firoj Alam, Giovanni Da San Martino, Maram Hasanain, RN Nandi, D Azizov, and P Panayotov. 2023. Overview of the clef-2023 checkthat! lab task 4 on factuality of reporting of news media. *Working Notes of CLEF*.

Preslav Nakov and Giovanni Da San Martino. 2021. Fake news, disinformation, propaganda, and media bias. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 4862–4865.

Preslav Nakov, Giovanni Da San Martino, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, and Thomas Mandl. 2021a. The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In *Advances in Information Retrieval – European Conference on IR Research*, pages 639–649.

Preslav Nakov, Giovanni Da San Martino, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, and Thomas Mandl. 2021b. The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In *Advances in Information Retrieval – 43rd European Conference on IR Research*, volume 12657 of *ECIR '21*, pages 639–649.

Wei Peng, Sue Lim, and Jingbo Meng. 2023. Persuasive strategies in online health misinformation: a systematic review. *Information, Communication & Society*, 26(11):2131–2148.

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 the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361.

Xiaoye Qu, Yingjie Gu, Qingrong Xia, Zechang Li, Zhefeng Wang, and Baoxing Huai. 2023. A survey on arabic named entity recognition: Past, recent advances, and future trends. *arXiv preprint arXiv:2302.03512*.

Adharsh Raj and Manash Pratim Goswami. 2020. Is fake news spreading more rapidly than covid-19 in india. *Journal of Content, Community and Communication*, 11(10):208–220.

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

Qiang Zhang, Aldo Lipani, Shangsong Liang, and Emine Yilmaz. 2019. Reply-aided detection of misinformation via bayesian deep learning. In *The world wide web conference*, pages 2333–2343.