Misinformation within Spotify Podcasts

Omar Jiménez
ojimnez3@gatech.edu
Georgia Institute of Technology
USA

Jonathan Leo
jpleo122@gatech.edu
Georgia Institute of Technology
USA

Abhijeet Tomar
atomar39@gatech.edu
Georgia Institute of Technology
USA

ABSTRACT

We propose a Natural Language Processing (NLP)-based approach to identify potentially misinformative claims in podcast transcripts. The proposed system uses a sentence-level analysis of podcast transcripts and claims from a fact-checked database by encoding these into a latent space, which is then explored to search for pairs with high semantic similarity. This approach resulted in SpotiFact, a dataset of more than 27 million similar pairs from the Spotify Podcast Dataset compiled to link stance (whether in agreement or not), out of which we manually label a subset of 1,200 claims from the top 3000 claims with highest similarity scores. We find that although most matched pairs are unlikely to be directly related, results suggest facts and misinformation spread at roughly the same rate. In addition, we build a set of machine learning models that are able to automatically detect misinformation in podcast transcripts, out of which a SentenceBERT + Linear Classifier performs best with an AUC score of 0.67. We run predictions on the rest of our dataset and results suggest misinformation spreads at a higher rate than factual claims within Spotify podcasts. We release the source code of this work at https://github.com/jpleo122/spotify-misinformation-cs-8803-DSN.

ACM Reference Format:
Omar Jiménez, Jonathan Leo, and Abhijeet Tomar. 2021. Misinformation within Spotify Podcasts. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years, misinformation and fake news detection on social networks have emerged as areas of interest due to the increasing amount of information consumption from online sources [12, 15, 16, 20]. A significant amount of work so far has been devoted to the study of misinformation on social media platforms (e.g., Facebook, Twitter), as well as the development of robust machine-learned systems to predict whether a given piece of information is true or false [11, 21, 23].

On the other hand, one could argue there are other regions within the social networks domain that have not received as much interest. A problem that appears to be unexplored, for example, is misinformation detection on podcasts. Podcasts are episodic series of audio/video files usually uploaded on streaming platforms (e.g., Spotify, Apple Podcasts) which are becoming increasingly popular sources for diverse types of news content. Although some podcast streaming platforms may share similarities with social media platforms, the problem of misinformation detection presents some unique challenges that require the development of novel approaches.

Limited Features: Social media platforms provide many ways for users to directly interact with content (e.g., likes, comments), which is not the case for most traditional podcast streaming platforms. This presents a challenge in the context of misinformation detection, where models often use abstract representations of these interactions as complimentary inputs (features) to generate their predictions. In addition to podcast transcripts and audio files, only a few additional resources may be available: episode title, summary, and other ones which may not be as valuable for models.

Heterogeneous Content: Usually, misinformation on social media occurs as *homogeneous* content (e.g., a fake news article, a post discussing a specific conspiracy theory). On the other hand, podcasts are well known for being a highly diverse type of content where creators often discuss a variety of topics (most of which are also not worthy of fact-checking). This is challenging in the context of targeting misinformation since it is unclear *where* to search for it.

Domain: Models for misinformation and fake news detection are typically representative of text data. Podcasts, on the other hand, are representative of speech data. This is also challenging since researchers have to rely on AI-generated transcripts as it is prohibitively expensive to transcribe hours upon hours of content.

Present Work: In this work, we propose an NLP-based approach to identify potentially misinformative claims in podcast transcripts. Overall, our main contributions are:

- **Novel problem definition:** To the best of our knowledge, we are the first to investigate the problem of misinformation detection on podcast streaming platforms.
- **Dataset:** We create SpotiFact, a dataset of more than 27 million similar pairs compiled to link stance (whether in agreement or not), out of which we manually label a subset of 1,200 claims from the top 3000 claims with highest similarity scores.
- **Discovery:** Our results suggest that facts and misinformation spread at roughly the same rate within Spotify podcasts.
- **Model:** We build a set of machine learning models that are able to automatically detect potentially misinformative claims within the podcast transcripts reasonably well.

2 RELATED WORK

2.1 Misinformation on Audiovisual Content

Recent works have examined misinformation on audiovisual platforms (e.g., Youtube) [6, 8, 19]. Although the concept of podcasts

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

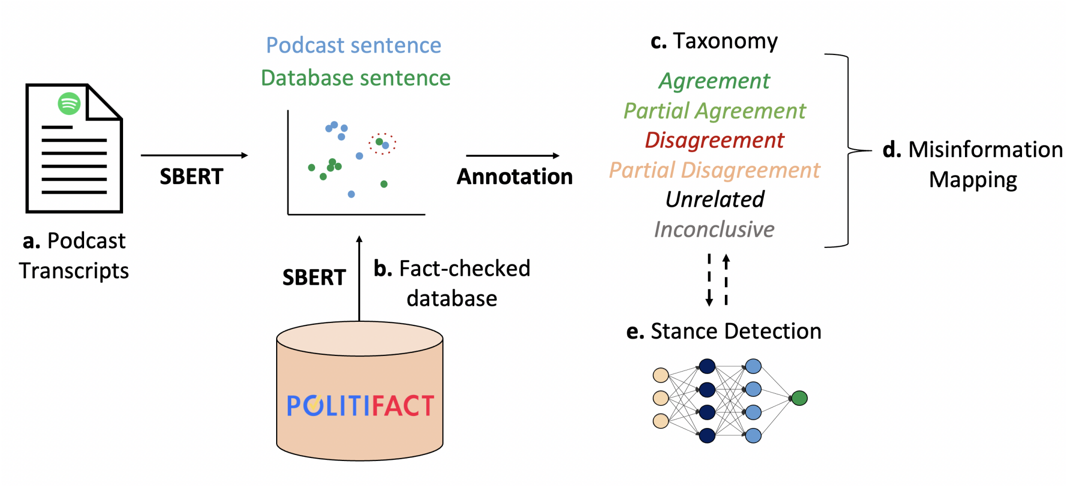

Figure 1: Proposed workflow. (a-b) Perform sentence-level embeddings of podcast transcripts and fact-checked database claims. We then explore the latent space and search for pairs with high semantic similarity and examine their stance, and use these labels to train a classifier (c, e). Finally, we perform logical operations to obtain misinformation labels using the ground-truth labels from Politifact (d).

has evolved over time and is now considered audiovisual content as well, most podcast episodes on traditional streaming platforms are purely audio-based. More importantly, all of the previously mentioned approaches rely on extracted features (e.g., comments, replies) that are not available on traditional podcast streaming platforms. We address this gap by proposing an approach that only uses podcast transcripts to identify potentially misinformative claims.

2.2 Misinformation Detection on Social Media

Misinformation detection (MID) in social networks is an active area of interest for researchers [7]. Many studies have been devoted to the development of models for MID on traditional social media platforms [1, 9, 14, 26]. However, to the best of our knowledge, no work has focused so far on identifying misinformation on podcast streaming platforms, which is the gap we address in this work.

2.3 Claim Verification

Claim verification is often recognized as a recognising textual entailment (RTE) or a natural language inference (NLI) task within the NLP domain [25]. In the case of models for RTE tasks, most of these were developed over a decade ago and hence do not exhibit desirable performance [22, 24]. State-of-the-art models on NLI datasets, on the other hand, have shown remarkable results in recent years [3, 13, 17]. Other efforts aim to explore claim verification trough the development of fully automated, end-to-end fact-checking platforms [5, 10]. The common issue with these models is that it is unlikely that they would be able to accurately capture the intricacies of speech data and the distinctive language use of the podcast domain, which are fundamentally different than text data. In this work we create SpotiFact, a novel dataset for claim verification containing transcribed speech data for over 27 million similar pairs of sentences which can be used to fine-tune these models.

3 DATASETS

3.1 Spotify Podcast Dataset

In this work we evaluate the proposed method with the Spotify Podcast Dataset [2], which one of the authors was provided access to by Spotify upon request. The dataset comprises podcast audio files along with AI-generated transcripts of over 100,000 Spotify podcast episodes published between January 1, 2019 and March 1, 2020. Episodes are randomly sampled from 18,376 podcasts with an average episode length of 5,700 transcribed words, 19.7 words per sentence and a vocabulary size of 562,318. The dataset also contains limited metadata related to creator-provided descriptions, geographic origins, RSS links, etc.

3.2 Fact-Checked Database

We crawled the *politifact.com* website to gather a database of fact-checked claims with their associated ground-truth labels. The collected database comprises 20287 claims spanning from May 2, 2007 to March 1, 2021 (in accordance with the timeframe of the Spotify Podcast Dataset) with an average of 17.9 words per claim and a vocabulary size of 43,358. We include fact-checked claims before the timeframe spanning the Spotify Podcast Dataset since, intuitively, it may occur that many of the claims contained within the transcripts were examined a long time before they were made.

4 METHODOLOGY

In this section, we describe our proposed workflow for misinformation targeting within Spotify Podcasts and the adopted taxonomy for stance annotation of the matched pairs dataset.

Sentence-Level Embeddings: Our workflow embeds all transcripts from the Spotify Podcast Dataset at a sentence-level, as well as all extracted claims from the *politifact.com* website. We use Sentence-BERT to compute embeddings, which has recently achieved remarkable results reducing the computational cost of semantic similarity search tasks by several orders of magnitude in comparison to traditional transformer models [18]. Obtained embeddings are of fixed length (384-dimensional vectors). In this work we use cosine similarity as the measure to compare obtained sentence embeddings.

Extracting Context: We note that isolated claims rarely contain sufficient information to enable appropriate labeling under the adopted stance annotation schema (which is described below). To this end, we also extract the two preceding and subsequent sentences of matched podcast sentences to provide additional context to the benefit of the annotators during the labeling process.

Taxonomy: We adopt a comprehensive 6-way stance annotation schema to properly label our dataset, which is described below with provided examples for each category. Note that although extracted context for podcast claims during the labeling process might be longer, we only show part of it here for brevity.

1. Agreement: Pairs of claims which are completely in harmony or accordance in opinion or feeling; although they may exhibit slight differences in syntax. The podcast transcript claim and the fact-checked claim are equivalent.

> **Fact-checked claim:** *Marijuana is a Schedule I drug, "which you understand means that you can't do any research about it."*
> **Podcast claim:** *" ... cannabis is scheduled so as a schedule 1 substance it basically what that means is that cannabis has quote no medicinal value, which means there's no reason to do research"*

2. Partial Agreement: Matched pairs where the podcast transcript shares the same stance as the fact-checked claim, but does not reference all the necessary components of the latter to be considered in full agreement (e.g., slight difference between numbers of relevant statistics).

> **Fact-checked claim:** *"Top CEOs (are) making 300 times the average worker."*
> **Podcast claim:** *"That CEOs making an exorbitant amount of money more than the employees is"*

In this case, the podcast statement agrees with the fact-checked claim, but does not specifically mention CEOs to be making *300 times* more than employees, hence *partial* agreement.

3. Disagreement: Pairs of claims which are opposite in opinion or feeling. The podcast transcript claim and the fact-checked claim are discussing the same topic but exhibit differing stances.

> **Fact-checked claim:** *"You've been tremendously deceived" by people who say the Earth is not flat."*
> **Podcast claim:** *"... it's very easy to prove that the Earth is not flat."*

4. Partial Disagreement: Matched pairs where the podcast transcript shares an opposite stance as the fact-checked claim, but does not reference all the necessary components of the latter to be considered in full disagreement.

> **Fact-checked claim:** *"During Obama's first five years as president, black unemployment increased 42 percent. During Reagan's presidency, black unemployment dropped 20 percent."*
> **Podcast claim:** *"What was the black unemployment rate when President Obama took office in 2008? Anybody Bueller Bueller? It was 16.8% Do you know that the first time when the unemployment rate for black people fell below 10% for the first time since the recession, you know, what year that was 2015, you know, wasn't it?. Office Donald Trump when Obama left office the black unemployment rate was 7.8%."*

The above pairs clearly disagree regarding the claim about black unemployment during Obama's presidency: one is saying it increased, the other it was reduced. However, the podcast claim does not mention anything about unemployment during Reagan's presidency, hence *partial* disagreement.

5. Unrelated: Two pairs are labeled as unrelated if they share no direct relation to each other.

> **Fact-checked claim:** *"Bill Gates admits his COVID-19 vaccine might kill nearly 1 million people."*
> **Podcast claim:** *"You know, Bill Gates is spending billions of dollars to provide vaccinations in Africa."*

This pair of claims exhibits high semantic similarity and they are both related to Bill Gates. However, they are clearly discussing different topics.

6. Inconclusive: Pairs of claims which might be related, but require additional context or information to enable a proper label under one of the previous categories.

> **Fact-checked claim:** *"Says she is a registered nurse."*
> **Podcast claim:** *"She's a registered nurse."*

The above pairs are almost semantically identical. However, clearly, it is difficult to know in advance if both claims are referring to the same person. Hence, we label this pair as *inconclusive*.

Misinformation Mapping: Using the labeled dataset and the ground truth labels (as obtained from *politifact.com*), we can determine whether a given claim is factual or misinformative using basic logic operations. The detailed mapping is shown below on Table 1. It is worth noting that in this work we reduce the fine-grained categorization used by Politifact into truth values (true, false and half-true), and adopt the following definition of misinformation: false, inaccurate, or misleading information.

Table 1: Misinformation mapping

Politifact ground-truth	Stance	Mapping
True Mostly True	Agreement Partial Agreement Disagreement Partial Disagreement	True Potentially True Misinformation Potential Misinformation
Barely-True, Mostly-False False, Pants-on-fire	Agreement Partial Agreement Disagreement Partial Disagreement	Misinformation Potential Misinformation True Potentially True
Half-True	Agreement Partial Agreement Disagreement Partial Disagreement	Misinformation Potential Misinformation Potential Misinformation* Potential Misinformation*

Table 2: Performance comparison for stance detection

Model	Accuracy	AUC
Frozen SentenceBERT + Linear Classifier	**0.668**	**0.669**
Frozen BERT-Large + Linear Classifier	0.668	0.525
Fine-tuned BERT-Large + Linear Classifier	0.668	0.496

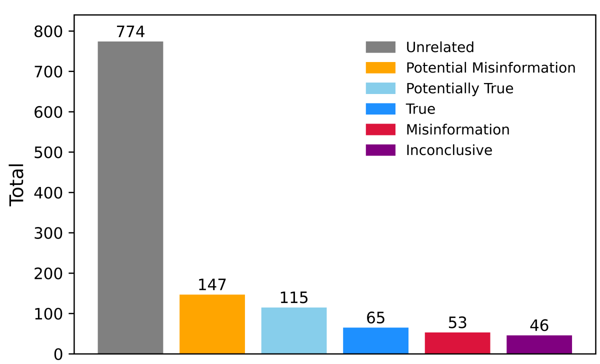

Figure 2: Results from manually labeled dataset after corresponding mappings.

5 RESULTS AND EXPERIMENTS

In this section, we examine the performance of the proposed workflow by conducting relevant experiments. Specifically, we aim to answer the following research questions:

- **RQ1:** Does misinformation spread occur in podcasts? How much in comparison to factual claims?
- **RQ2:** Are there any intrinsic characteristics of misinformation within podcasts (e.g., channels that exhibit an abnormal amount of false claims, podcast length)?
- **RQ3:** Can we effectively predict whether a podcast contains misinformation?

5.1 SpotiFact Dataset

Annotation Process: The annotation process is carried out by all three authors of this paper. Each annotator labels an overlapping set of 600 matched pairs randomly sampled from the dataset, as well as 200 additional individual ones for each annotator (total of 1,200 pairs). At least two out of the three annotators agreed on 579 of the overlapping 600 claims (95.8%), and the remaining 21 pairs where all annotators disagreed were reconsidered after discussion. Overall, the annotation process resulted in a Cohen's Kappa score of 0.54, indicating a moderate inter-rater agreement that is similar to other works on stance annotation in this topic [4].

Labeled Dataset: We summarize the results of our labeling process in Fig. 2. Recall that we use a 6-way labeling process in which we categorize a pair of matched claims as one of the following: *agreement*, *partial agreement*, *disagreement*, *partial disagreement*, *unrelated*, or *inconclusive*. From Figure 2, it is evident that most of our labeled data comprises unrelated claims. In general, we find that these are pairs with high semantic similarity but do not share a direct relation with each other (see examples from Section 4). On the other hand, the rest of our labeled dataset may provide revealing results. We find a total of 65 *true* and 53 *misinformative* claims, as well as 115 *potentially true* and 147 *potentially misinformative* claims. These results suggest facts and misinformation spread at roughly the same rate within our labeled dataset. Finally, we also find 46 *inconclusive* claims.

5.2 Stance Classifier

Benchmarking: We use the hand-annotated dataset to train a text-based machine learning classifier to label matched pairs using the same 6-way stance schema previously described. We experimented with three different models, where each model takes as input a fact checked claim and a paired podcast claim plus it's surrounding context. Both of these are embedded using a different sentence embedding technique and fed through a linear classifier to predict their stance agreement. The three embedding layers are a frozen

SentenceBERT model, a frozen BERT-Large model, and a BERT-Large model that is fine-tuned with the classifier layer [3, 18].

Performance of the models was measured using accuracy and ROC AUC, as shown in Table 2. In the table, we note that accuracy performance is very similar for all embedding methods. On the other hand, SentenceBERT embedding features perform significantly better than BERT and BERT-Large embeddings with an AUC score of 0.67. We also note that AUC provides a much more appropriate measure of model performance due to the class imbalance in our dataset and determine SentenceBERT + Linear Classifier to be the model that performs best.

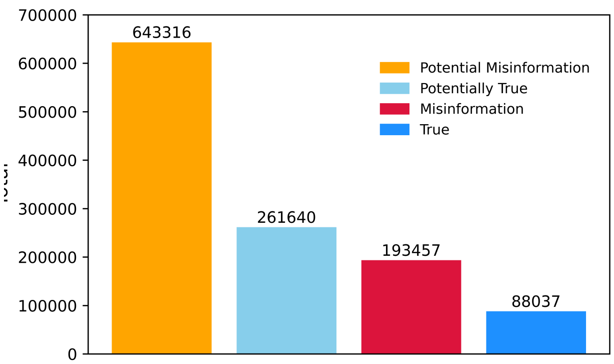

Figure 3: Results from running the SentenceBERT + Linear Classifier model on half of the matched pairs (13 million in total). Unrelated and inconclusive claims are excluded for clarity.

Model Predictions: Using the best performing model (SentenceBERT + Linear Classifier), we predict the stance agreement between half of the rest of the matched pairs in the dataset. This results in about 13 million pairs that took around 7 hrs to predict. Results are summarized in Fig. 3. We find that around 70.5% of the relevant pairs (i.e., not labeled as unrelated or inconclusive) comprise potentially false claims, suggesting that misinformation spreads at a higher rate in our dataset.

6 LIMITATIONS AND FUTURE WORK

In this work, we rely on AI-generated transcripts as inputs to our proposed workflow. The manual annotation process revealed that often these are not as accurate as desired, which inherently affects the quality of our assigned labels. Other limitations include the necessary restriction of not introducing domain knowledge to the labeling process and that the presented work is purely a text-based approach. Future work seeks to leverage the proposed workflow with respective episode audio files, which may prove useful in this context. Other potential directions are the inclusion of additional fact-checked databases, fine-tuning transformer models for stance detection with labeled data from SpotiFact, as well as adopting finer-grained definitions of what comprise *claims*. Finally, future work also seeks an end-to-end implementation of our proposed system, providing a useful way to flag potentially false claims on Spotify podcasts that can be further examined by human annotators and/or domain experts.

7 CONCLUSION

To our best knowledge, this paper is the first attempt to explore misinformation detection in traditional podcast streaming platforms. The proposed workflow resulted in a dataset of 27,666,306 matched claims, out of which we manually label a random subset of 1,200 claims from the top 3000 highest ones using a 6-way stance-based taxonomy (*agreement*, *disagreement*, *partial agreement/disagreement*, *unrelated*, *inconclusive*). Results suggest that, although most potential candidates extracted from our workflow are unlikely to contain misinformation, facts and misinformation spread at roughly the same rate within Spotify podcasts. Moreover, predictions on the rest of our dataset suggest misinformation spreads at a higher rate than factual claims. This simple, yet effective methodology can be used for investigating other podcast streaming platforms for misinformation detection. We hope that this work will attract more attention from the community towards developing accurate misinformation detection models for podcast streaming platforms.

8 CONTRIBUTIONS

Omar Jiménez contributed the general project design, code for data analysis and data collection (crawling), annotator, manuscript. Jonathan Leo contributed modifications to project design, code for data analysis and model training, experiments, code refactoring, annotator. Abhijeet Tomar contributed modifications to project design, code for data analysis, experiments, annotator.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (May 2017), 211–36. https://doi.org/10.1257/jep.31.2.211
[2] Ann Clifton, Sravana Reddy, Yongze Yu, Aasish Pappu, Rezvaneh Rezapour, Hamed Bonab, Maria Eskevich, Gareth Jones, Jussi Karlgren, Ben Carterette, et al. 2020. 100,000 podcasts: A spoken English document corpus. In *Proceedings of the 28th International Conference on Computational Linguistics*. 5903–5917.
[3] 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* (2018).
[4] Andreas Hanselowski, Christian Stab, Claudia Schulz, Zile Li, and Iryna Gurevych. 2019. A richly annotated corpus for different tasks in automated fact-checking. *arXiv preprint arXiv:1911.01214* (2019).
[5] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1803–1812.
[6] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
[7] Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, and Guandong Xu. 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining* 10, 1 (2020), 1–20.
[8] Raj Jagtap, Abhinav Kumar, Rahul Goel, Shakshi Sharma, Rajesh Sharma, and Clint P. George. 2021. Misinformation Detection on YouTube Using Video Captions. arXiv:2107.00941 [cs.LG]
[9] Suchita Jain, Vanya Sharma, and Rishabh Kaushal. 2016. Towards automated real-time detection of misinformation on Twitter. In *2016 International conference on advances in computing, communications and informatics (ICACCI)*. IEEE, 2015–2020.
[10] Lev Konstantinovskiy, Oliver Price, Mevan Babakar, and Arkaitz Zubiaga. 2021. Toward Automated Factchecking: Developing an Annotation Schema and Benchmark for Consistent Automated Claim Detection. *Digital Threats: Research and*

Practice 2, 2 (2021), 1–16.

[11] Srijan Kumar and Neil Shah. 2018. False Information on Web and Social Media: A Survey. *CoRR* abs/1804.08559 (2018). arXiv:1804.08559 http://arxiv.org/abs/1804.08559

[12] Thai Le, Suhang Wang, and Dongwon Lee. 2020. MALCOM: Generating Malicious Comments to Attack Neural Fake News Detection Models. *CoRR* abs/2009.01048 (2020). arXiv:2009.01048 https://arxiv.org/abs/2009.01048

[13] 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:1907.11692 [cs.CL]

[14] Alexandre Maros, Jussara Almeida, Fabrício Benevenuto, and Marisa Vasconcelos. 2020. Analyzing the Use of Audio Messages in WhatsApp Groups. https://doi.org/10.1145/3366423.3380070

[15] Kai Nakamura, Sharon Levy, and William Yang Wang. 2019. r/Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. *CoRR* abs/1911.03854 (2019). arXiv:1911.03854 http://arxiv.org/abs/1911.03854

[16] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. *CoRR* abs/2008.07939 (2020). arXiv:2008.07939 https://arxiv.org/abs/2008.07939

[17] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683* (2019).

[18] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[19] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. NLP-based feature extraction for the detection of COVID-19 misinformation videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

[20] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. DEFEND: Explainable Fake News Detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 395–405. https://doi.org/10.1145/3292500.3330935

[21] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *CoRR* abs/1708.01967 (2017). arXiv:1708.01967 http://arxiv.org/abs/1708.01967

[22] Marta Tatu and Dan Moldovan. 2007. COGEX at RTE3. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing* (Prague, Czech Republic) *(RTE '07)*. Association for Computational Linguistics, USA, 22–27.

[23] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559

[24] Fabio Massimo Zanzotto, Marco Pennacchiotti, and Alessandro Moschitti. 2009. A machine learning approach to textual entailment recognition. *Natural Language Engineering* 15, 4 (2009), 551–582.

[25] Xia Zeng, Amani S Abumansour, and Arkaitz Zubiaga. 2021. Automated fact-checking: A survey. *Language and Linguistics Compass* 15, 10 (2021), e12438.

[26] 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*. 2333–2343.