🦕 STANCEOSAURUS: Classifying Stance Towards Multicultural Misinformation

Jonathan Zheng, Ashutosh Baheti, Tarek Naous, Wei Xu, Alan Ritter
School of Interactive Computing
Georgia Institute of Technology
{jonathanqzheng, abaheti3, tareknaous}@gatech.edu; {wei.xu, alan.ritter}@cc.gatech.edu

Abstract

We present *Stanceosaurus*, a new corpus of 28,033 tweets in English, Hindi, and Arabic annotated with stance towards 251 misinformation claims. As far as we are aware, it is the largest corpus annotated with stance towards misinformation claims. The claims in Stanceosaurus originate from 15 fact-checking sources that cover diverse geographical regions and cultures. Unlike existing stance datasets, we introduce a more fine-grained 5-class labeling strategy with additional subcategories to distinguish implicit stance. Pre-trained transformer-based stance classifiers that are fine-tuned on our corpus show good generalization on unseen claims and regional claims from countries outside the training data. Cross-lingual experiments demonstrate Stanceosaurus' capability of training multi-lingual models, achieving 53.1 F1 on Hindi and 50.4 F1 on Arabic without any target-language fine-tuning. Finally, we show how a domain adaptation method can be used to improve performance on Stanceosaurus using additional RumourEval-2019 data. We make Stanceosaurus publicly available to the research community and hope it will encourage further work on misinformation identification across languages and cultures.[1]

1 Introduction

The prevalence of misinformation on online social media has become an increasingly severe societal problem. A key language technology, which has the potential to help content moderators identify rapidly-spreading misinformation, is the automatic identification of both affective and epistemic stance (Jaffe, 2009; Zuczkowski et al., 2017) towards false claims. Progress on the problem of stance identification has largely been driven by the availability of annotated corpora, such as RumourEval (Derczynski et al., 2017; Gorrell et al., 2019). However,

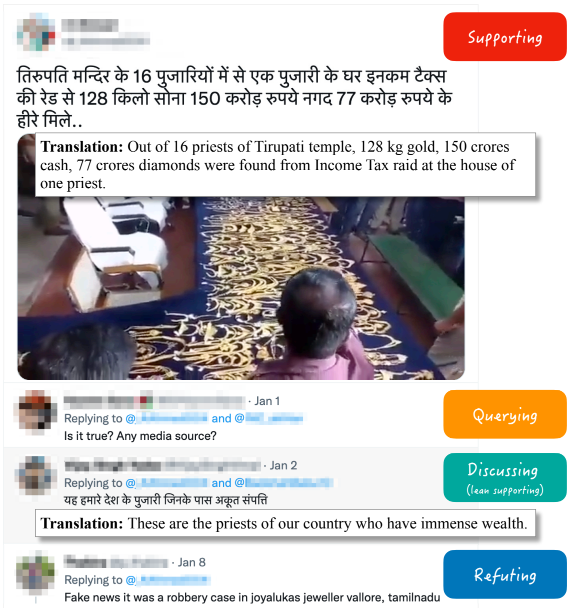

Figure 1: Example Hindi and English tweets in Stanceosaurus with stance towards the claim *"Raid at Tirupati temple priest's house, 128 kg gold found"*.

existing corpora mostly focus on misinformation spreading within western countries.

In this paper, we present *Stanceosaurus*, a diverse and high-quality corpus that builds on the best design choices made in previous misinformation corpora, including RumourEval-2019 and CovidLies. Stanceosaurus covers more diverse topics, geographic regions, and cultures than prior work. It includes 28,033 tweets in English, Hindi, and Arabic that are manually annotated for stance (see Figure 1) towards 251 misinformation claims, collected from 15 independent fact-checking websites that cover India, Singapore, Australia, New Zealand, Canada, the United States, Europe, and the Arab World (the Levantine, Gulf, Northwest African regions, and Egypt). To the best of our knowledge, Stanceosaurus is the largest and most diverse annotated stance dataset to date.

Through extensive experiments, we demonstrate that Stanceosaurus can support the fine-grained

[1]Our code and data are available at https://tinyurl.com/stanceosaurus

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 2132–2151
December 7-11, 2022 ©2022 Association for Computational Linguistics

Dataset	Target	Number/Range of Topics
SemEval-2016 (Mohammad et al., 2016)	Subject	6 political topics (e.g., *atheism*, *feminist movement*)
SRQ (Villa-Cox et al., 2020)	Subject	4 political topics & events (e.g., *general terms*, *student marches*)
Catalonia (Zotova et al., 2020)	Subject	1 topic (i.e., *Catalonia independence*)
COVID (Glandt et al., 2021)	Subject	4 topic related to Covid-19 (e.g., *stay at home orders*)
Multi-target (Sobhani et al., 2017)	Entity	3 pairs of candidates in 2016 US election
WTWT (Conforti et al., 2020)	Event	5 merger and acquisition events
RumourEval (Gorrell et al., 2019)	Tweet	8 news events + rumors about natural disasters
Rumor-has-it (Qazvinian et al., 2011)	Claim	5 rumors (e.g., *Sarah Palin getting divorced?*)
CovidLies (Hossain et al., 2020)	Claim	86 pieces of COVID-19 misinformation
Stanceosaurus (this work)	Claim	**251** claims over a diverse set of global and regional topics

Table 1: Summary of Twitter stance classification datasets. Stanceosaurus covers more claims from a broader range of topics and geographical regions than prior Twitter stance datasets.

classification of explicit and implicit stances, as well as zero-shot cross-lingual stance identification. In addition, we introduce and experiment with class-balanced focal loss (Cui et al., 2019) to alleviate the class imbalance issue, which is a well-known challenge in automatic stance detection (Zubiaga et al., 2016; Baly et al., 2018). Similar to other corpora that are labeled with stance towards messages or claims, Stanceosaurus reflects the natural distribution of stance observed *in the wild*, with comparatively few examples labeled as Supporting or Refuting (see label distributions in Table 4). We show that fine-tuning BERTweet$_{large}$ with class-balanced focal loss (Cui et al., 2019) can achieve 66.8 F1 for 3-way stance classification and 61.0 F1 for the finer-grained 5-way stances for English. With zero-shot transfer learning, we achieve 50.4 and 53.1 F1 for Hindi and Arabic, respectively, in a 5-way classification. Lastly, we show it is possible to train a single model to achieve better performance on Stanceosaurus' test set via additional fine-tuning on RumourEval (Gorrell et al., 2019), using a variation of EasyAdapt (Daumé III, 2007; Bai et al., 2021) designed for pre-trained Transformers, even though these two corpora have significant differences.

2 Related Work

Stance Classification Datasets. Given the importance of studying misinformation spreading on Twitter and the open access to its data, there are many stance classification datasets consisting of annotated tweets. However, existing datasets are largely restricted to a limited range and a number of topics — see Table 1 for a summary.[2] Note that many of these datasets are considering stance toward an entity or topic (e.g., *Bitcoin*), whereas we focus on more specific full-sentence claims (e.g., *Bitcoin is legal in Malaysia*), which provides flexibility to cover more diverse topics in our work.

The closet prior efforts to ours are RumourEval-2019 (Gorrell et al., 2019) and CovidLies (Hossain et al., 2020). RumourEval-2019 (Gorrell et al., 2019) contains annotations on whether a reply tweet in a conversation thread is supporting, denying, querying, or commenting on the rumour mentioned in the source tweet. However, RumourEval covers only eight major news events (e.g., *Charlie Hebdo shooting*) plus additional rumors about natural disasters. The CovidLies dataset (Hossain et al., 2020) annotates a 3-way stance (Agree, Disagree, Neutral) towards 86 pieces of COVID-19-specific misinformation, using BERTScore (Zhang et al., 2020) to find potentially relevant tweets. As the authors of CovidLies (Hossain et al., 2020) have noted, relying on BERTScore (i.e., a semantic similarity measurement) biases the data collection towards more supporting and less refuting tweets.

Besides Twitter, stance classification has also been studied for other types of data. For example, the Perspectrum dataset (Chen et al., 2019) was constructed using debate forum data. Emergent (Ferreira and Vlachos, 2016) and AraStance (Alhindi et al., 2021) consist of English and Arabic news articles annotated with stance, respectively.

Fact Checking Datasets. Related to but different from stance classification, fact-checking (aka rumour verification) as an NLP task primarily focuses on the assessment of claims being true or false. There exist several fact-checking datasets, such as FEVER (Thorne et al., 2018) and MultiFC (Augenstein et al., 2019) for English, and X-Fact (Gupta and Srikumar, 2021) for 25 non-English languages. These datasets consist of claims extracted from Wikipedia or fact-checking sites, which are

[2] See also the excellent survey by Hardalov et al. (2021c). Given space limitations, we highlight only the most relevant work.

Source	Country & Regions	Lang	#Claims	#Tweets	Irr.	Sup.	Ref.	Dis.	Que.
Snopes	USA (80%), INT'L (16.7%), Other (3.3%)	en	30	3197	1051	428	229	1447	42
Poynter	Europe (5%), INT'L (90%), Other (5%)	en	20	2197	949	274	97	844	33
FullFact	UK (30%), INT'L (55%), Other (15%)	en	20	2379	806	300	179	1057	37
AFP Fact Check CAN	Canada (55%), INT'L (30%), Other (15%)	en	20	2078	746	252	130	910	40
AAP Fact Check	Australia (10%), INT'L (65%), Other (25%)	en	20	2302	739	374	136	1019	34
AFP Fact Check NZ	New Zealand (15%), INT'L (75%), Other (10%)	en	20	2227	879	194	81	1044	29
Blackdotresearch	Singapore (30%), INT'L (55%), Other (15%)	en	20	2307	842	248	113	1076	28
Factly	India (45%), INT'L (55%)	en	20	1979	889	190	117	734	49
Politifact	USA (20%), INT'L (35%), Other (45%)	en	20	2041	984	289	8	753	7
Alt News	India (90.4%), INT'L (4.8%), Other (4.8%)	hi	21	1730	550	489	172	500	19
Aajtak	India (67%), Other (33%)	hi	9	806	456	110	40	193	7
Hindi Newschecker	India (56%), Other (44%)	hi	9	781	195	313	46	219	8
MISBAR	Arab World (58.3%), INT'L (8.3%), Other (33.4%)	ar	12	2283	454	514	203	1031	81
Fatabyyano	Arab World (28.5%), INT'L (57.1%), Other (14.4%)	ar	7	986	234	163	49	522	18
Maharat Fact-o-meter	INT'L (100%)	ar	3	740	224	132	55	316	13
Total	Regional (57.4%), INT'L (42.6%)		251	28033	9998	4270	1655	11665	445

Table 2: Fact-checking sources included in our Stanceosaurus corpus. The most common regions are listed. **Stance** – breakdown of tweets into 5 main categories in relation to each claim: Irrelevant, Supporting, Refuting, Discussing, and Querying. **Country & Regions** – home country of the source and the distribution of claims regarding home country, regional, and international matters. Other refers to claims in countries other than the primary countries covered by the source (e.g., Snopes claims about India).

labeled for veracity.

Automatic Stance Classification. Many prior efforts have developed methods for automatic stance classification, which have progressed from feature-based approaches (Qazvinian et al., 2011; Lukasik et al., 2015; Ferreira and Vlachos, 2016; Zeng et al., 2016; Aker et al., 2017; Riedel et al., 2017; Zhang et al., 2018; Ghanem et al., 2018; Lukasik et al., 2019; Li et al., 2019a) to neural approaches (Kochkina et al., 2017; Chen et al., 2017; Veyseh et al., 2017; Bhatt et al., 2018; Hanselowski et al., 2018; Poddar et al., 2018; Umer et al., 2020), then to fine-tuning of pre-trained models (Ghosh et al., 2019; Fajcik et al., 2019; Matero et al., 2021). Researchers (Zubiaga et al., 2016) have noted the class imbalance issue in stance classification and subsequently chose Macro F1 as the main evaluation metric. To deal with imbalanced data, previous works have used methods such as per-label weights (García Lozano et al., 2017; Ghanem et al., 2019), oversampling underrepresented examples (Singh et al., 2017), retrieving additional examples from external datasets (Yang et al., 2019), or adjusting prediction thresholds over class label probabilities (Li and Scarton, 2020). With the availability of many small-scale stance datasets, other works attempted weakly supervised (Kumar, 2020; Yang et al., 2022), semi-supervised methods (Giasemidis et al., 2020), or multi-task models (Kochkina et al., 2018; Ma et al., 2018; Li et al., 2019b; Wei et al., 2019; Kumar and Carley, 2019; Fang et al., 2019; Cheng et al., 2020; Yu et al., 2020; Khandelwal, 2021). A few efforts have also looked at transferring knowledge from larger datasets to smaller datasets (Xu et al., 2018; Hardalov et al., 2021a; Schiller et al., 2021) and languages with less data (Mohtarami et al., 2019; Zotova et al., 2020; Hardalov et al., 2021b). Stanceosaurus (this work) is one of the largest and most diverse stance classification datasets to date, enabling the study of cross-lingual transfer for stance classification towards misinformation claims.

3 The Stanceosaurus Corpus

Our corpus consists of social media posts manually annotated for stance toward claims from multiple fact-checking websites across the world. We carefully designed the data collection and annotation scheme to ensure better quality and coverage, improving upon prior work.

3.1 Collecting Fact-checked Claims

To ensure multicultural representation, we obtain fact-checked claims from both Western and non-Western sources (Table 2). We choose nine well-known fact-checking websites in English, three in Hindi, and three in Arabic.[3] We randomly select claims from each source posted between 5/17/2012 and 02/28/2022 that have sparked discussion on Twitter. In total, we have 251 claims in our corpus, of which 144 are considered regional based on manual inspection (see column **Country & Regions** in

[3] Fact-checking sources are selected from Wikipedia, Poynter's International Fact-Checking Network, as well as those in X-Fact (Gupta and Srikumar, 2021).

Table 2). For example, the claims *"Finland is promoting a 4 day work week"* and *"Burning Ghee will produce Oxygen"*[4] are both considered regional, one explicitly and one implicitly; whereas the claim *"Bees use acoustic levitation to fly"* is considered international. The claims in Stanceosaurus range from news, health, and science to politics (e.g., *"Sonu Sood promises to support Hamas/Palestine"*), conspiracy theories, history, and urban myths (e.g., *"The pyramids of Giza were built by slaves"*). We present all 251 claims in Appendix D.

3.2 Retrieving Conversations around Claims

For better coverage of diverse topics, we invested substantial effort in creating customized queries with varied keywords and time ranges for each claim to retrieve tweets. We also trace the entire reply chain in both directions, so Stanceosaurus includes relevant tweets that may not contain the keywords.

Curated Search Queries. We retrieve tweets by keyword search, which we believe is the most effective approach given the constraints of Twitter's APIs. To ensure the coverage and quality of our dataset, we manually curated and iteratively refined search queries for each claim, utilizing advanced search operators to restrict the relevant time period and language. We expand search queries with synonyms (e.g., *"jab"* for *"vaccine"*) and lexical variations whenever possible; the latter is particularly helpful for including different Arabic dialects. See Appendix A for example queries. We collect tweets from different time periods for different claims (e.g., a two-week range for timely events and a max range from 7/3/2008 to 5/9/2022 for historic myths).

Context from URLs and Reply Chains. Individual tweets retrieved by search do not capture the contextual aspects of stance, which can be very important as misinformation often spreads in multi-turn conversations on social media. Therefore, we also collect the parent tweets (i.e., the tweet that a search retrieved tweet is replying to) and the entire reply chains if available. Additional details are presented in Appendix C.

3.3 Annotating Stance Towards Claims

We employ a fine-grained annotation scheme that supports 5-way and 3-way stance classification.

5-way Stance Categories. We define stance detection as a five-way classification task, including irrelevant tweets in addition to the four stance classes used in prior works (Schiller et al., 2021; Gorrell et al., 2018), as follows:

- **Irrelevant** – unrelated to the claim;
- **Supporting** – explicitly affirms the claim is true or provides verifying evidence;
- **Refuting** – explicitly asserts the claim is false or presents evidence to disprove the claim;
- **Discussing** – provide neutral information on the context or veracity of the claim;
- **Querying** – questions the veracity of the claim.

See Figure 1 and Appendix B.1 for examples of different stances, shown with the reply chain details.

Subcategories and 3-way Stance Classification. Although some tweets may be neutral towards a claim, they can still show an indirect bias. For example, the tweet *"Fauci: No Concern About Number of People Testing Positive After COVID-19 Vaccine."* in response to the claim *"The COVID-19 Vaccine has magnets or will make your body magnetic"* discusses the vaccine rollout, while it can be viewed as implicitly supporting the claim regarding the lack of vaccine safety. We thus further annotate the Discussing tweets for their leanings as three subcategories: Discussing$_{support}$ (44.6%), Discussing$_{refute}$ (25.7%), and Discussing$_{other}$ (29.7%). This not only enables fine-grained classification but also makes our Stanceosaurus corpus flexible enough to support the 3-way (Supporting, Refuting, Other)[5] setup used in other prior work.

Data Annotation. We hired four native speakers for English, two for Hindi, and two for Arabic to annotate the tweets with stance. English annotators are all from the U.S., and non-English annotators grew up in the respective countries or regions of the claims being collected. All of the annotators have a college-level education. We designed detailed guidelines (see Appendix B.2) and held training sessions to assist our annotators. For each claim, the annotators are reasonably familiar with the topic because they are asked to read and learn about the subject matter before annotating. Cohen's Kappa (κ) between the annotators is summarized in Table 3, showing substantial agreement (Artstein and Poesio, 2008) for all languages.

[4]Ghee is a type of clarified butter, commonly used in cuisines from the Indian subcontinent.

[5]By merging (1) Discussing$_{support}$ with Supporting; (2) Discussing$_{refute}$ with Refuting; (3) Discussing$_{other}$, Irrelevant, and Querying together into Other.

#Tweets	Lang	5-class κ	3-class κ
20,707	en	0.624	0.670
3,317	hi	0.673	0.742
4,009	ar	0.773	0.729

Table 3: Inter-annotator agreement calculated based on 5-class and 3-class stances.

Disagreements often occur over challenging cases. For example, *"Evergreen ship stuck in the Suez Canal - interesting call sign"* is supporting the conspiracy theory *"Hillary Clinton is trafficking children aboard the Evergreen Ship"*, with the connection being that the call sign of the ship is *"H3RC"*, which coincidentally overlaps with Hillary's initials. The disagreements were resolved by a third adjudicator for Hindi, and through discussions between the annotators for English and Arabic. Interestingly, the Hindi subset of Stanceosaurus exhibits some forms of code-switching in 28.2% of instances, including some replies written in English, while 6.3% of the Arabic data exhibited code-switching. A subset of 200 tweets randomly sampled from the Arabic data was further labeled for language variations, which contains 62.5% Modern Standard Arabic (MSA), 35.5% dialects, 0.5% Arabizi, and 1.5% in the form of emojis or mentions.

3.4 Comparison to RumourEval

Although our annotation design is comparable to RumourEval (Gorrell et al., 2019), in that both annotate the stance of Twitter threads towards rumorous claims, there are a few important differences: (1) RumourEval limits their rumorous claims primarily to 8 major news events plus additional natural disaster events, whereas we use a much larger and more diverse sample of claims originating from multicultural news outlets. (2) RumourEval, unlike our dataset, does not explicitly provide the claims. Rather, the first tweet of the thread is used to represent both the claim and the stance in RumourEval. (3) We label discussing subcategories that capture indirect bias towards a claim (see §3.3). (4) RumourEval excludes irrelevant tweets, limiting its generalizability. For a direct comparison, we present the corpus statistics of Stanceosaurus and RumourEval-2019[6] in Table 4, and further test

Stance	Stanceosaurus #train	Stanceosaurus #test	Stanceosaurus #dev	RumourEval #train	RumourEval #test	RumourEval #dev
Irrelevant	4928	1674	1283	—	—	—
Supporting	1462	592	495	925	157	102
Refuting	598	270	222	378	101	82
Discussing	4941	2160	1783	3507	1405	1174
+Other	949	532	366	—	—	—
+Supporting	2440	1082	780	—	—	—
+Refuting	1552	546	637	—	—	—
Querying	201	54	44	395	93	120
Total	12130	4750	3827	5205	1756	1478

Table 4: (Left) Number of tweets in the English subset of Stanceosaurus. The train/dev/test sets consist of 112/44/34 separate claims, respectively. (Right) Statistics of RumourEval-2019 (Gorrell et al., 2019) after we reconstruct the data from message IDs.

classification models on both datasets in §5.3.

4 Automatic Stance Detection

We design multiple automatic stance identification experiments to test the generalization capabilities of models trained on Stanceosaurus. First, we establish the baseline performance of predicting stance towards unseen claim using fine-tuned Transformer models in §5.1 and experiment with the class-balanced focal loss for addressing the imbalanced class distribution. We present zero-short cross-lingual experiments in §5.2, where multilingual models are trained on English tweets and evaluated on the Hindi and Arabic tweets. Furthermore, we demonstrate that a simple domain adaptation method can help improve performance on Stanceosaurus using additional RumourEval data in §5.3. Finally, we show that models trained only on International claims subset can extrapolate well to regional claims from individual countries in §5.4.

4.1 Baseline Models

We experiment with fine-tuning methods using BERT (Devlin et al., 2019) and BERTweet (Nguyen et al., 2020). The latter is a RoBERTa-based (Liu et al., 2019) model pre-trained on Twitter data.[7] Stance identification is modeled as sentence-pair classification, using special tokens to format the input as "[CLS] claim [SEP] text", where "text" is a tweet concatenated with its context (parent tweet and any extracted HTML titles – see §3.2). We found that incorporating context generally helps stance classification for reply tweets (see ablation

[6]As RumourEval distributes only message IDs, we reconstructed the dataset by retrieving all the available posts from Twitter and Reddit, with a loss of a small portion of data that has been deleted on the social media platform (120 out of 1876 instances in the test set; 12 and 7 instances in the train/dev).

[7]The *base* size of the BERTweet model is trained on 850M English tweets streamed from 01/2012 to 08/2019. The *large* size is trained with additional 23M tweets that are related to COVID-19.

study in Appendix B.3). We use standard cross-entropy loss in all baselines.

4.2 Class-balanced Focal Loss (CB_{foc})

The imbalanced class problem has been identified as a major challenge in automatic stance classification (Li and Scarton, 2020), since fewer messages exhibit Supporting or Refuting stances in the wild (see Table 4). To alleviate this issue, prior work has used weighted cross-entropy loss (Fajcik et al., 2019). We experiment with weighted cross-entropy loss and Class-Balanced Focal loss (Cui et al., 2019; Baheti et al., 2021), which has shown promising results in computer vision research recently, as an alternative.

We use $\hat{s} = (z_0, z_1, z_2, z_3, z_4)$ to represent the unnormalized scores assigned by the model for five stance classes C = {Irrelevant, Discussing, Supporting, Refuting, Querying}. The class-balanced focal loss is then defined as:

$$\mathrm{CB_{foc}}(\hat{s}, y) = -\underbrace{\frac{1-\beta}{1-\beta^{n_y}}}_{\text{reweighting}} \underbrace{\sum_{m \in C} (1-p_m)^{\gamma} \log(p_m)}_{\text{focal loss}} .$$

y is the gold stance label, n_y is the number of instances with the label y, and $p_m = sigmoid(z'_m)$, where:

$$z'_m = \begin{cases} z_m & m = y \\ -z_m & \text{otherwise} \end{cases}$$

Focal loss employs the expression $(1 - p_m)^{\gamma}$ to reduce the relative loss for well classified examples (Lin et al., 2017). The reweighting term lowers the impact of class imbalance on the loss. In our experiments, hyperparameters β and γ are tuned between [0.1, 1) and [0.1, 1.1], respectively, based on the performance on the dev set.

4.3 Implementation Details

We replace usernames and URLs with special tokens, truncate or pad the input to a sequence length of 256 as in BERT and BERTweet.[8] All models were trained for 10 epochs and optimized with the Adam optimizer. Learning rates were selected among $\{1e^{-5}, 3e^{-5}, 5e^{-5}, 7e^{-5}, 9e^{-5}\}$. The train batch size was set to 8. For all test set evaluations, we select the best checkpoint that achieves the highest Macro F1 on the development set.

[8]We use the maximum sequence length of 128 tokens for BERTweet$_{base}$.

5 Experiments and Results

We report average results over five random seeds primarily by Macro F1, which has been used as the standard metric in stance classification since the arguably more important stances (i.e., Refuting and Supporting) only consist of a small portion of data.

5.1 Stance Detection for Unseen Claims

For this experiment, we split the English data based on claims into train, dev, and test set (see the left side of Table 4). We evaluate all models on the 5-way stance classification of tweets towards claims that are unseen during training. As shown in Table 5, the best model is BERTweet$_{large}$, which achieves 60.2 F1 when trained with standard and weighted cross-entropy loss and 61.0 F1 with class-balanced focal loss. We see some alleviation of the data imbalance issue in the per-label analysis in Table 6, which shows improved F1 using class-balanced focal loss for the two least frequent labels, Refuting and Querying.

As mentioned in §3.3, Stanceosaurus can also support 3-way stance classification by merging Discussing$_{support}$ and Discussing$_{refute}$ tweets with Supporting and Refuting, respectively. We present the results from BERTweet$_{large}$ for this experiment in Table 6. Interestingly, the label F1 for Refuting decreases in the 3-way classification, compared to the 5-way setup. It suggests that identifying the indirect leaning for Discussing$_{refute}$ tweets makes the task harder. Meanwhile, the higher F1 scores for Supporting and Other labels indicate that our classifier is good at detecting tweets that propagate

Model	Stanceosaurus (unseen claims) Precision	Recall	F1
BERT$_{base}$+ CE	51.1$\pm$1.1	50.5$\pm$2.0	50.4$\pm$1.6
+ weighted CE	50.5$\pm$1.9	52.7$\pm$1.1	51.3$\pm$1.3
+ CB_{foc}	50.6$\pm$1.3	55.7$\pm$2.1	52.5$\pm$1.0
BERT$_{large}$+ CE	54.3$\pm$0.8	53.0$\pm$0.6	53.6$\pm$0.6
+ weighted CE	53.8$\pm$1.3	53.8$\pm$1.2	53.6$\pm$1.0
+ CB_{foc}	53.9$\pm$1.2	53.7$\pm$1.1	53.6$\pm$0.5
BERTweet$_{base}$+ CE	53.1$\pm$1.2	52.2$\pm$1.6	52.3$\pm$1.0
+ weighted CE	51.8$\pm$1.0	55.2$\pm$1.4	53.1$\pm$0.7
+ CB_{foc}	51.3$\pm$0.6	56.8$\pm$0.6	53.5$\pm$0.3
BERTweet$_{large}$+ CE	60.6$\pm$2.0	60.2$\pm$1.0	60.2$\pm$1.1
+ weighted CE	**60.8**$\pm$1.6	60.2$\pm$1.0	60.2$\pm$0.5
+ CB_{foc}	59.8$\pm$1.3	**62.8**$\pm$1.5	**61.0**$\pm$0.8

Table 5: 5-way stance classification results for unseen claims in Stanceosaurus (mean$\pm$standard deviation across runs of five random seeds). Class-balanced focal loss (CB_{foc}) outperforms standard and weighted cross-entropy loss (CE, weighted CE).

Stance Class		#test	Cross-Entropy Loss Precision	Recall	F1	Weighted Cross-Entropy Loss Precision	Recall	F1	Class-balanced Focal Loss Precision	Recall	F1
5-Class	Supporting	592	**59.9**±2.1	61.5±2.2	**60.6**±1.0	57.2±0.9	60.6±2.3	58.8±1.6	57.6±1.0	**63.8**±1.3	60.5±1.0
	Refuting	270	60.6±6.9	57.4±1.9	58.7±2.3	60.6±3.0	**61.6**±4.6	60.9±1.1	**60.9**±2.2	**61.6**±3.2	**61.1**±1.0
	Discussing	2160	66.4±0.7	63.5±2.7	**64.9**±1.3	65.5±1.1	**64.1**±2.9	64.7±1.0	**67.0**±1.1	60.0±1.8	63.2±0.8
	Querying	54	43.7±7.2	42.6±5.9	42.5±3.7	**47.5**±6.8	41.1±4.3	43.6±2.7	42.3±5.7	**51.5**±6.6	**45.8**±2.9
	Irrelevant	1674	72.4±2.4	76.0±2.3	**74.1**±0.9	**73.2**±2.3	73.4±5.0	73.2±1.5	71.1±0.9	**77.4**±2.8	**74.1**±0.9
	All	3912	60.6±2.0	60.2±1.0	60.2±1.1	**60.8**±1.6	60.2±1.0	60.2±0.5	59.8±1.3	**62.8**±1.5	**61.0**±0.8
3-Class	Supporting	1674	66.9±1.6	68.1±1.3	67.5±1.3	68.9±3.1	**68.2**±4.6	**68.3**±1.3	**70.1**±1.5	65.0±2.9	67.4±1.0
	Refuting	816	55.2±1.8	51.9±4.5	53.3±1.6	**56.0**±2.5	52.2±2.4	53.9±1.5	54.5±3.0	**58.5**±5.1	**56.2**±0.8
	Other	2260	75.9±1.3	76.4±1.2	76.1±0.5	75.9±1.9	**77.9**±2.9	76.8±0.6	**76.2**±1.6	**77.9**±1.6	**77.0**±0.2
	All	3912	66.0±0.4	65.5±1.1	65.6±0.7	66.9±0.9	66.1±1.0	66.4±0.9	**66.9**±0.5	**67.1**±0.9	**66.8**±0.4

Table 6: Per-label comparison of BERTweet$_{large}$, when fine-tuned with cross-entropy, weighted cross-entropy loss, and class-balanced focal loss, both for 3-class and 5-class stance detection on our corpus. Weighted cross-entropy and class-balanced focal loss improves F1 score overall, and in particular for the least frequent stance of Refuting.

misinformation, even when some of them do not assert a stance explicitly.

5.2 Zero-Shot Cross-Lingual Transfer

Truly multicultural stance identification requires models that are capable of operating across languages. To demonstrate the feasibility of identifying the stance towards misinformation claims in a zero-shot cross-lingual setting, when no training data in the target language is available, we fine-tune models on Stanceosaurus' English training set and use all the annotated Hindi/Arabic data as the test set. We experiment with both multilingual BERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020). Because we assume no training data is available for the target language, all hyperparameters are tuned on the English dev set. Full results of our 5-class cross-lingual experiments are presented in Table 7. When trained with class-balanced focal loss, XLM-RoBERTa$_{large}$ achieves 53.1 Macro F1 for Hindi and 50.4 for Arabic, notably outperforming models trained with cross-entropy loss.

Model	Stanceosaurus (English → Hindi) Precision	Recall	F1
mBERT$_{base}$ + CE	52.1±2.9	39.4±2.0	40.8±2.5
+ weighted CE	55.0± 4.2	42.4± 1.4	44.3±1.8
+ CB$_{\text{foc}}$	53.0±3.4	44.1±1.7	45.3±1.5
XLM-R$_{base}$ + CE	53.2±0.1	42.6±2.1	44.3±1.9
+ weighted CE	50.3± 3.2	44.4± 1.9	44.6±1.5
+ CB$_{\text{foc}}$	52.8±2.1	46.5±0.7	47.4±0.9
XLM-R$_{large}$ + CE	55.7±3.5	49.0±1.8	49.9±1.6
+ weighted CE	**57.5**± 1.3	51.1± 0.9	52.5± 1.0
+ CB$_{\text{foc}}$	57.4±2.1	**51.5**±1.3	**53.1**±1.6
	Stanceosaurus (English → Arabic)		
mBERT$_{base}$ + CE	44.8±4.0	40.1±2.5	40.0±2.0
+ weighted CE	44.1± 3.3	40.7± 1.6	39.7±1.7
+ CB$_{\text{foc}}$	46.1±2.6	44.7±1.1	43.1±0.2
XLM-R$_{base}$ + CE	47.6±1.8	41.9±2.1	42.6±2.2
+ weighted CE	46.1± 2.0	47.9± 2.5	46.1±2.1
+ CB$_{\text{foc}}$	45.8±1.7	50.0±2.2	46.4±1.6
XLM-R$_{large}$ + CE	51.4±2.7	49.2±3.4	47.7±2.3
+ weighted CE	49.6± 1.3	49.7± 1.7	48.2±1.4
+ CB$_{\text{foc}}$	**51.9**±2.0	**52.2**±2.6	**50.4**±0.5

Table 7: Cross-lingual experiments where the models are trained on the English part of Stanceosaurus and evaluated on the Hindi/Arabic data. Models trained with class-balanced focal loss (CB$_{\text{foc}}$) outperforms those trained with standard and weighted cross-entropy loss (CE) with higher Macro F1 and lower variance.

5.3 Combining Stanceosaurus + RumourEval

Because Stanceosaurus follows a similar labeling scheme as existing stance corpora, such as RumourEval (Gorrell et al., 2018), this raises a natural question: is it possible to achieve better performance by combining the two datasets?

We first confirm that fine-tuning BERTweet$_{large}$ with class-balanced focal loss is also the best performing model on RumourEval-2019's original 4-class evaluation setup, outperforming the weighted cross-entropy loss used in BUT-FIT (Fajcik et al., 2019),[9] as shown in Table 8. To evaluate cross-dataset performance, we then convert both Stanceosaurus and RumourEval-2019 into 3-way stances to minimize the differences between their annotation schemes. RumourEval is converted by collapsing Discussing and Querying instances into

[9]BUT-FIT (Fajcik et al., 2019) is one of the state-of-the-art methods on RumourEval-2019, following closely (0.2% lower Macro F1) behind the winning system BLCU_NLP (Yang et al., 2019). BLCU_NLP achieved a Macro F1 score of 0.62 and used specialized features but is not open-sourced.

Model	RumourEval-2019 Precision	Recall	F1
BERT$_{large}$ + CE	**66.8**±3.5	51.8±2.3	56.0±1.8
+ weighted CE	61.8±4.5	**56.7**±3.8	56.7±3.9
+ CB$_{foc}$	62.5±6.0	54.6±1.9	**57.5**±2.8
BERTweet$_{large}$ + CE	68.6±5.0	**62.4**±1.3	64.0±1.5
+ weighted CE	68.4±4.3	62.1±2.5	63.0±2.9
+ CB$_{foc}$	**74.4**±3.9	61.8±1.8	**65.7**±1.4

Table 8: Results on RumourEval-2019 that compare different models trained with class-balanced focal loss (CB$_{foc}$), standard and weighted cross-entropy losses.

the Other category. When merging the datasets, we upsample the RumourEval dataset to twice its size to counteract the imbalance between the two datasets. Table 9 shows models trained on in-domain data achieve higher performance than the naive merging of the two datasets for training.

To close this performance gap, we adopt the **EasyAdapt** (Daumé III, 2007; Bai et al., 2021) method to fine-tune BERTweet$_{large}$ on the combination of RumourEval and Stanceosaurus. EasyAdapt creates three identical copies of the contextualized representations of the input, which are concatenated and fed into a linear layer before softmax classification. The parameters in the linear layer that correspond to the first and third copies are updated when training on Stanceosaurus, while others are zeroed out; the parameters that correspond to the second and third copies are updated when training on RumourEval. This enables the model to encode representations that are specific to each dataset and domain-independent parameters that can transfer between the two datasets. BERTweet$_{large}$ with EasyAdapt achieves 67.4 Macro F1 for Stanceosaurus and 65.8 Macro F1 for RumourEval, outperforming the in-domain model performance for Stanceosaurus and matching the in-domain model performance of RumourEval.

5.4 Stance Detection for Unseen Countries

The English dataset comprises 97 international and 93 regional claims. We test BERTweet's ability to generalize toward regional claims by training on international claims. Specifically, we create a new train-test-dev split, with 10740/5701/4896 datapoints spread around 97/43/42 claims. Table 10 shows the results stratified by source. Performance on the regional data varies widely between sources. Poynter and AFP Fact Check New Zealand, two sources with the most international data, have the best F1s at 63.0 and 63.5 respectively.

Train \ Test	Stanceosaurus Prec.	Rec.	F1	RumourEval Prec.	Rec.	F1
Stanceosaurus	66.9	**67.1**	66.8	44.8	43.8	41.2
RumourEval	39.8	43.6	40.6	**79.6**	59.7	65.7
Combined	66.6	66.0	66.2	61.1	63.4	60.6
EasyAdapt	**68.3**	67.0	**67.4**	74.4	**62.6**	**65.8**

Table 9: Cross-domain experiments on Stanceosaurus and RumourEval. We fine-tune BERTweet$_{large}$ using class-balanced focal loss. Performance drops significantly when training on one dataset and testing on the other. However, with **EasyAdapt** (Daumé III, 2007; Bai et al., 2021), we attain a single model that achieves best performance on Stanceosaurus while being on-par with in-domain RumourEval model performance.

Fact Check Source	#test	Precision	Recall	F1
AAP Fact Check	452	50.1	39.4	40.6
AFP Fact Check CAN	824	71.5	54.7	58.7
AFP Fact Check NZ	224	64.2	63.7	63.5
Blackdotresearch	516	65.7	62.0	60.4
Factly	447	59.4	68.2	62.5
FullFact	474	57.0	55.4	55.8
Poynter	118	73.2	61.3	63.0
Politifact	614	57.7	53.7	51.8
Snopes	1402	61.4	52.0	54.4
All	5071	62.9	54.3	57.1

Table 10: Results on Unseen Countries experiment. BERTweet$_{large}$ finetuned on class-balanced focal loss is trained on international claims and evaluated on regional claims, stratified by fact-checking source. The model achieves an aggregate F1 that is somewhat lower than its counterpart in the Unseen Claims experiment.

6 Conclusion

We introduce Stanceosaurus, a new corpus of 28,033 social media messages annotated with stance towards 251 misinformation claims originating from 15 multicultural fact-checking sources. To the best of our knowledge, Stanceosaurus is the largest stance dataset yet. Stanceosaurus contains consistent annotations across claims and languages, and stance classifier models trained on our dataset can perform well on unseen claims and languages. Our experiments demonstrate that class-balanced focal loss consistently improves upon cross-entropy loss in addressing the stance label-imbalance issue. Furthermore, the domain adaptation experiments with EasyAdapt show it is possible to utilize RumourEval data to achieve even better performance on Stanceosaurus despite significant differences in their data collection strategies. Our work represents a step towards the development of accurate models that can track the spread of misinformation online across diverse languages and cultures.

Limitations

We currently use manually curated search queries for collecting tweets related to misinformation claims in Stanceosaurus. While we tried our best to include relevant keywords and their synonyms in the search queries, it still requires careful manual effort and may not be exhaustive in finding all relevant tweets related to the claim. Furthermore, it is non-trivial to extend such queries to new claims and languages. Future work could look at automatically generating these queries using a few-shot shot in-context demonstrations with large language models (Brown et al., 2020).

We collect the Stanceosaurus dataset with all the human resources available to us for three languages. We leave annotations for more languages for future work. We will also release our detailed data annotation guideline and invite other researchers to extend our work to set a standard benchmark for stance classification.

There are also potential biases in the claims that reflect the biases of content moderators from the fact-checking sources. We made our best effort to identify a list of fact-checking sources based on Wikipedia and pre-existing datasets used in the NLP community to collect claims from different countries and languages. We randomly sample claims from these sources and, since we are constructing a Twitter-based dataset, we are only able to include claims that have been discussed on Twitter. If the claim is unpopular on Twitter, we cannot sufficient data for annotation. Following Twitter's Developer Agreement and Policy, we release our dataset freely for academic research and include the full set of claims in the Appendix of the paper for readers to examine the potential biases in our dataset more conveniently.

Although the class-balanced focal loss improves stance classification in data imbalanced settings, our models are still far from perfect. We do not use user-specific, temporal, and network features as additional context which has been shown to improve prediction performance (Aldayel and Magdy, 2019; Lukasik et al., 2016).

Broader Impact and Ethical Considerations

We will release our dataset under Twitter Developer Agreement,[10] which grants permissions for academic researchers to share Tweet IDs and User IDs for non-commercial research purposes, as of October 1st. 2022.

Our datasets and models are developed for research purposes and may contain unknown biases towards certain demographic groups or individuals (Sap et al., 2019). Further investigation into systematic biases should be conducted before deployment in a production environment.

Social media companies currently struggle with content moderation in non-Western countries.[11] We hope Stanceosaurus will help stimulate more public research that can help shed light on how to inhibit the spread of dangerous misinformation across languages and cultures.

Acknowledgments

We thank three anonymous reviewers for their helpful comments. We also thank Chao Jiang for providing his codebase; Chase Perry, Mohamed Ghanem, Angana Borah, Rucha Sathe, Andrew Duffy, and Kenneth Koepcke for their help with data annotation. This research is supported in part by NSF awards IIS-2144493 and IIS-2052498, in addition to ODNI and IARPA via the BETTER program (contract 19051600004). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of NSF, ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Ahmet Aker, Leon Derczynski, and Kalina Bontcheva. 2017. Simple Open Stance Classification for Rumour Analysis. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, 2017*, pages 31–39, Varna, Bulgaria. INCOMA Ltd.

Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proceedings of the ACM Human Computer Interaction*, 3.

Tariq Alhindi, Amal Alabdulkarim, Ali Alshehri, Muhammad Abdul-Mageed, and Preslav Nakov. 2021. AraStance: A Multi-Country and Multi-Domain Dataset of Arabic Stance Detection for Fact

[10] https://developer.twitter.com/en/developer-terms/agreement-and-policy

[11] https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/

Checking. In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, 2021*, pages 57–65, Online. Association for Computational Linguistics.

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4):555–596.

Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. 2019. MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking of Claims. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4685–4697, Hong Kong, China. Association for Computational Linguistics.

Ashutosh Baheti, Maarten Sap, Alan Ritter, and Mark Riedl. 2021. Just Say No: Analyzing the Stance of Neural Dialogue Generation in Offensive Contexts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4846–4862, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Fan Bai, Alan Ritter, and Wei Xu. 2021. Pre-train or annotate? domain adaptation with a constrained budget. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5002–5015, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, and Preslav Nakov. 2018. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 21–27.

Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. 2018. Combining Neural, Statistical and External Features for Fake News Stance Identification. In *International World Wide Web Conference (Companion Volume)*, pages 1353–1357.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Sihao Chen, Daniel Khashabi, Wenpeng Yin, Chris Callison-Burch, and Dan Roth. 2019. Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 542–557, Minneapolis, Minnesota. Association for Computational Linguistics.

Yi-Chin Chen, Zhao-Yang Liu, and Hung-Yu Kao. 2017. IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification. In *Proceedings of the 11th International Workshop on Semantic Evaluation 2017*, pages 465–469, Vancouver, Canada. Association for Computational Linguistics.

Mingxi Cheng, Shahin Nazarian, and Paul Bogdan. 2020. Vroc: Variational autoencoder-aided multi-task rumor classifier based on text. In *Proceedings of the web conference 2020*, pages 2892–2898.

Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. Will-They-Won't-They: A Very Large Dataset for Stance Detection on Twitter. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1715–1724, Online. 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.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge J. Belongie. 2019. Class-Balanced Loss Based on Effective Number of Samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277.

Hal Daumé III. 2007. Frustratingly Easy Domain Adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263.

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 2017*, pages 69–76, Vancouver, Canada. Association for Computational Linguistics.

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*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Martin Fajcik, Pavel Smrz, and Lukas Burget. 2019. BUT-FIT at SemEval-2019 Task 7: Determining the Rumour Stance with Pre-Trained Deep Bidirectional Transformers. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1097–1104, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Wei Fang, Moin Nadeem, Mitra Mohtarami, and James Glass. 2019. Neural Multi-Task Learning for Stance Prediction. In *Proceedings of the Second Workshop on Fact Extraction and VERification*, pages 13–19, Hong Kong, China. Association for Computational Linguistics.

William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1163–1168, San Diego, California. Association for Computational Linguistics.

Marianela García Lozano, Hanna Lilja, Edward Tjörnhammar, and Maja Karasalo. 2017. Mama Edha at SemEval-2017 Task 8: Stance Classification with CNN and Rules. In *Proceedings of the 11th International Workshop on Semantic Evaluation 2017*, pages 481–485, Vancouver, Canada. Association for Computational Linguistics.

Bilal Ghanem, Alessandra Teresa Cignarella, Cristina Bosco, Paolo Rosso, and Francisco Manuel Rangel Pardo. 2019. UPV-28-UNITO at SemEval-2019 Task 7: Exploiting Post's Nesting and Syntax Information for Rumor Stance Classification. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1125–1131, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2018. Stance Detection in Fake News A Combined Feature Representation. In *Proceedings of the First Workshop on Fact Extraction and VERification*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Shalmoli Ghosh, Prajwal Singhania, Siddharth Singh, Koustav Rudra, and Saptarshi Ghosh. 2019. Stance Detection in Web and Social Media: A Comparative Study. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 75–87, Cham. Springer International Publishing.

Georgios Giasemidis, Nikolaos Kaplis, Ioannis Agrafiotis, and Jason R. C. Nurse. 2020. A Semi-Supervised Approach to Message Stance Classification. *IEEE Transactions on Knowledge and Data Engineering*, 32(1):1–11.

Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance Detection in COVID-19 Tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 1596–1611, Online. Association for Computational Linguistics.

Genevieve Gorrell, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. RumourEval 2019: Determining Rumour Veracity and Support for Rumours.

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. SemEval-2019 task 7: RumourEval, Determining Rumour Veracity and Support for Rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Ashim Gupta and Vivek Srikumar. 2021. X-Fact: A New Benchmark Dataset for Multilingual Fact Checking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 675–682, Online. Association for Computational Linguistics.

Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 2018. A Retrospective Analysis of the Fake News Challenge Stance-Detection Task. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1859–1874, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021a. Cross-Domain Label-Adaptive Stance Detection. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9011–9028, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021b. Few-shot cross-lingual stance detection with sentiment-based pre-training. *arXiv preprint arXiv:2109.06050*.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021c. A Survey on Stance Detection for Mis- and Disinformation Identification. *CoRR*, abs/2103.00242.

Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. 2020. COVIDLies: Detecting COVID-19 Misinformation on Social Media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.

Alexandra Jaffe. 2009. *Stance: Sociolinguistic Perspectives*. Oxford University Press.

Anant Khandelwal. 2021. Fine-Tune Longformer for Jointly Predicting Rumor Stance and Veracity. In *8th ACM IKDD CODS and 26th COMAD*, CODS COMAD 2021, page 10–19, New York, NY, USA. Association for Computing Machinery.

Elena Kochkina, Maria Liakata, and Isabelle Augenstein. 2017. Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM. In *Proceedings of the 11th International Workshop on Semantic Evaluation 2017*, pages 475–480, Vancouver, Canada. Association for Computational Linguistics.

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, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sumeet Kumar. 2020. Social Media Analytics for Stance Mining A Multi-Modal Approach with Weak Supervision.

Sumeet Kumar and Kathleen Carley. 2019. Tree LSTMs with Convolution Units to Predict Stance and Rumor Veracity in Social Media Conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5047–5058, Florence, Italy. Association for Computational Linguistics.

Quanzhi Li, Qiong Zhang, and Luo Si. 2019a. eventAI at SemEval-2019 Task 7: Rumor Detection on Social Media by Exploiting Content, User Credibility and Propagation Information. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 855–859, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Quanzhi Li, Qiong Zhang, and Luo Si. 2019b. 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, Florence, Italy. Association for Computational Linguistics.

Yue Li and Carolina Scarton. 2020. Revisiting Rumour Stance Classification: Dealing with Imbalanced Data. In *Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media*, pages 38–44, Barcelona, Spain (Online). Association for Computational Linguistics.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. 2017. Focal Loss for Dense Object Detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2999–3007.

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. *CoRR*, abs/1907.11692.

Michal Lukasik, Kalina Bontcheva, Trevor Cohn, Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2019. Gaussian Processes for Rumour Stance Classification in Social Media. *ACM Trans. Inf. Syst.*, 37(2).

Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. 2015. Classifying Tweet Level Judgements of Rumours in Social Media. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2590–2595, Lisbon, Portugal. Association for Computational Linguistics.

Michal Lukasik, P. K. Srijith, Duy Vu, Kalina Bontcheva, Arkaitz Zubiaga, and Trevor Cohn. 2016. Hawkes Processes for Continuous Time Sequence Classification: an Application to Rumour Stance Classification in Twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 393–398, Berlin, Germany. Association for Computational Linguistics.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Detect Rumor and Stance Jointly by Neural Multi-task Learning. In *International World Wide Web Conference (Companion Volume)*, pages 585–593.

Matthew Matero, Nikita Soni, Niranjan Balasubramanian, and H. Andrew Schwartz. 2021. MeLT: Message-Level Transformer with Masked Document Representations as Pre-Training for Stance Detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2959–2966, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation 2016*, pages 31–41, San Diego, California. Association for Computational Linguistics.

Mitra Mohtarami, James Glass, and Preslav Nakov. 2019. Contrastive Language Adaptation for Cross-Lingual Stance Detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4442–4452, Hong Kong, China. Association for Computational Linguistics.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 9–14, Online. Association for Computational Linguistics.

Lahari Poddar, Wynne Hsu, Mong Li Lee, and Shruti Subramaniyam. 2018. Predicting Stances in Twitter Conversations for Detecting Veracity of Rumors: A Neural Approach. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 65–72.

Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying Misinformation in Microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *arXiv e-prints*, pages arXiv–1707.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678.

Benjamin Schiller, Johannes Daxenberger, and Iryna Gurevych. 2021. Stance detection benchmark: How robust is your stance detection? *KI-Künstliche Intelligenz*, pages 1–13.

Vikram Singh, Sunny Narayan, Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya. 2017. IITP at SemEval-2017 Task 8 : A Supervised Approach for Rumour Evaluation. In *Proceedings of the 11th International Workshop on Semantic Evaluation 2017*, pages 497–501, Vancouver, Canada. Association for Computational Linguistics.

Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2017. A Dataset for Multi-Target Stance Detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 551–557, Valencia, Spain. Association for Computational Linguistics.

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*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics.

Muhammad Umer, Zainab Imtiaz, Saleem Ullah, Arif Mehmood, Gyu Sang Choi, and Byung-Won On. 2020. Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). *IEEE Access*, 8:156695–156706.

Amir Pouran Ben Veyseh, Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2017. A Temporal Attentional Model for Rumor Stance Classification. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 2335–2338, New York, NY, USA. Association for Computing Machinery.

Ramon Villa-Cox, Sumeet Kumar, Matthew Babcock, and Kathleen M. Carley. 2020. Stance in Replies and Quotes (SRQ): A New Dataset For Learning Stance in Twitter Conversations.

Penghui Wei, Nan Xu, and Wenji Mao. 2019. Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4787–4798, Hong Kong, China. Association for Computational Linguistics.

Brian Xu, Mitra Mohtarami, and James R. Glass. 2018. Adversarial Domain Adaptation for Stance Detection. *32nd Conference on Neural Information Processing Systems 2018*.

Ruichao Yang, Jing Ma, Hongzhan Lin, and Wei Gao. 2022. A Weakly Supervised Propagation Model for Rumor Verification and Stance Detection with Multiple Instance Learning. *arXiv preprint arXiv:2204.02626*.

Ruoyao Yang, Wanying Xie, Chunhua Liu, and Dong Yu. 2019. BLCU_NLP at SemEval-2019 Task 7: An Inference Chain-based GPT Model for Rumour Evaluation. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1090–1096, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

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*, pages 1392–1401, Online. Association for Computational Linguistics.

Li Zeng, Kate Starbird, and Emma Spiro. 2016. # Unconfirmed: Classifying Rumor Stance in Crisis-Related Social Media Messages. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 747–750.

Qiang Zhang, Emine Yilmaz, and Shangsong Liang. 2018. Ranking-based Method for News Stance Detection. In *International World Wide Web Conference (Companion Volume)*, pages 41–42.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations*.

Elena Zotova, Rodrigo Agerri, Manuel Nuñez, and German Rigau. 2020. Multilingual Stance Detection in Tweets: The Catalonia Independence Corpus. In *Proceedings of the 12th Language Resources* *and Evaluation Conference*, pages 1368–1375, Marseille, France. European Language Resources Association.

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. 2016. Stance Classification in Rumours as a Sequential Task Exploiting the Tree Structure of Social Media Conversations. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers 2016*, pages 2438–2448, Osaka, Japan. The COLING 2016 Organizing Committee.

Andrzej Zuczkowski, Ramona Bongelli, and Ilaria Riccioni. 2017. *Epistemic stance in dialogue: knowing, unknowing, believing*, volume 29. John Benjamins Publishing Company.

A Customized Queries for Retrieving Tweets

We present example claims and their search queries from each of the three languages in Table 13.

B Stance Classification with Context

B.1 Annotation Example of Tweets in Reply Chain

Table 14 shows representative examples of different stances towards the claim *"The COVID-19 Vaccine will make your body magnetic"*. Note that some tweets are context-dependent (e.g., *"No that is not true"*); their stance can only be determined with appropriate context.

B.2 Guidelines for Tricky Annotation

We identified some common scenarios in our annotation which lead to annotation disagreements in our preliminary analysis of the data. We designed specific guidelines to improve annotation consistency, including:

- If the claim has a lot of information, we should focus on the core contentious part of the claim when judging the stance of the tweets.

- If the tweet is giving an analysis of the contentious event or talking about an adjacent event (regional) then it should be considered Discussing.

- If the tweet is just emojis, praise, or pleasant message (e.g., "*thank you*", "*good job sir*") towards a context tweet, consider it Discussing with the leaning inherited from the Stance of the context tweet.

- For querying, the tweet should be questioning the veracity of the claim and not any other question about the incident.

- If the main purpose of the tweet is gauging the people's opinions related to the claim then it is Discussing.

- If the tweet is posing a question with #fakenews or #factcheck but the URL asserts that the claim is fake then it should be judged Refuting. However, if the URL is also a question without a judgment then it should be considered Discussing.

- If a reply tweet is adding information/opinion on top of the context (assuming that the context tweet is true) then annotate Discussing with Leaning inherited from the context.

B.3 Importance of Considering Context

Stance that is realized in social media messages often depends on the context of a conversation, or links to external webpages, as discussed in §3.2. In this section, we evaluate the impact of context in the form of parent tweets and URL titles. To ablate context, we first organize tweets in the training data into reply chains. Next, we separate threads into *root tweets* that have no parent in the conversation thread and *reply tweets* that are written in response to another message. We fine-tune BERTweet$_{large}$ on (1) only root tweets, (2) only reply tweets, and (3) both root and reply tweets. We also measure the impact of training with and without context. We use standard cross-entropy loss for this comparison study, excluding the impact of hyperparameter choices in the focal loss, as the stance distribution differs between root and reply tweets.

The results in Table 11 demonstrate that root tweets, reply tweets, and context are complementary for achieving the best overall performance. The F1 score on root tweets is significantly higher than on reply tweets, indicating the difficulty to determine stance in extended conversations. Unsurprisingly, training only on root tweets achieves a higher 61.7 F1 on root tweets but a lower 35.4 F1 on reply tweets. For models trained only on reply tweets, including context improves performance on reply tweets but hurts performance on root tweets.

B.4 Unseen Fact-checking Sources

Since the claims in Stanceosaurus are collected from multicultural sources, we also test stance clas-

Train \ Test	Root Tweets (56.1%) Precision	Recall	F1	Reply Tweets (43.9%) Precision	Recall	F1	All Tweets Precision	Recall	F1
▷ root	**60.8**±1.2	63.0±1.8	**61.7**±1.3	35.1±1.2	40.4±2.1	35.4±1.5	52.4±0.4	56.1±1.4	53.6±0.9
▷ reply w/o context	53.7±3.2	42.1±4.2	43.9±4.7	33.1±1.5	35.3±2.0	33.2±1.3	44.7±2.9	39.7±3.4	40.5±3.3
▷ reply w/ context	49.0±3.2	37.3±0.7	38.6±1.5	42.6±3.3	41.3±2.4	41.6±2.5	47.7±3.7	39.1±1.4	41.0±2.2
▷ root + reply w/o context	60.2±1.4	63.2±1.7	61.4±0.8	35.7±0.9	42.1±1.1	36.9±0.9	52.1±1.1	56.6±1.2	53.8±0.7
▷ root + reply w/ context	60.3± 2.0	**63.5**± 1.8	61.5± 1.5	**56.7**± 4.0	**46.2**± 1.8	**49.5**± 2.0	**66.0**±0.4	**65.5**±1.1	**65.6**±0.7

Table 11: Ablation experiments to study the impact of context in 5-way stance classification. In particular, we split the Twitter threads within Stanceosaurus' training set into root tweets (those with no parent in the conversation thread) and reply tweets (tweets that are written in response to another message). In all experiments, we train BERTweet$_{large}$ using cross-entropy loss. Results suggest that predicting the stance of reply tweets is significantly harder than root tweets. Context improves the overall stance classification performance mainly by improving prediction on reply tweets.

Fact Check Source	Unseen Source #train*	#test	F1	Unseen Claims F1
AAP Fact Check	11135	477	58.7	**59.2**
AFP Fact Check CAN	10941	459	57.6	**59.6**
AFP Fact Check NZ	10554	482	54.5	**55.7**
Blackdotresearch	10699	517	57.3	**59.3**
Factly	10693	318	59.4	**61.4**
FullFact	10783	602	60.0	**61.8**
Politifact	10927	838	50.6	**60.0**
Poynter	10715	321	52.4	**55.3**
Snopes	10593	736	57.7	**58.0**
All	12130	4750	-	**61.0**

Table 12: Results of BERTweet$_{large}$ with the class-balanced focal loss on unseen fact-checking sources. For each source, we remove associated tweets from train/dev in Stanceosaurus' standard data split. Macro F1 scores are computed on a subset of the test set with tweets only from the unseen source. We also report the performance of the same model trained on full train/dev splits in Stanceosaurus with tweets from all sources. Performance is degraded when predicting stance on unseen sources, but not by a large margin.

sifier's performance towards claims found in fact-checking sources that are unseen in the training data. Specifically, we convert each fact-checking website in Stanceosaurus into an unseen source by creating a new data-split and removing its tweets from the train and dev sets. Then, a model trained on this restricted data is evaluated on the test tweets from the selected unseen source. For comparison, we also report the performance of the best model from the unseen claims experiment (§5.1 where claims from each source are split into train/dev/test) on these test tweets from the unseen source. For every unseen source, we train a BERTweet$_{large}$ stance classifier with class-balanced focal loss and report its results in Table 12. The models perform worse when the source is removed from training data, with Politifact showing the biggest drop in performance from 60.0 F1 to 50.6 F1. This highlights the importance of source-specific data in classifying misinformation claims.

C Additional Details on Conversation Threads

For each claim from English and Hindi sources, we randomly sample up to 150 tweets for annotation: max 50 tweets (average 50 for English and 48.1 for Hindi) retrieved from our queries, max 50 parent tweets (average 30.7 for English and 8.6 for Hindi), and max 50 children tweets (average 28.3 for English and 33.0 for Hindi) from reply chains. For Arabic, we annotated all the tweets (average 175.8 per claim) retrieved from the search and reply chain. Finally, we organize every tweet such that its immediate parent serves as the context. For tweets containing URLs, we also additionally include the HTML 'Title' tag extracted from the URL. About 40.5% of all tweets in our dataset have a parent tweet in context, while 19.5% of tweets have associated HTML titles.

D *Stanceosaurus* Claims

We provide the full set of English, Hindi, and Arabic claims, with English translations for Hindi and Arabic. Figure 2 provides all 22 Arabic claims, Table 15 provides all 39 Hindi claims, and Table 16 provides all 190 English claims.

Claim	Query
Easter is a celebration for the Mediterranean Goddess Ishtar	(easter ishtar) lang:en -filter:retweets
The false positive rate for a COVID-19 test is very high	((COVID OR coronavirus) AND false positive) lang:en -filter:retweets
पाकिस्तान में ग्रह युद्ध छिड़ गया है। पाक आर्मी और कराची पुलिस के बीच जबरदस्त फायरिंग शुरू हो गई है। *Translation:* In-house war has broken out in Pakistan. Heavy firing has started between Pak Army and Karachi Police.	(पाकिस्तान OR पाक) (युद्ध OR फायरिंग OR आर्मी) (कराची OR पुलिस) since:2020-09-01 until:2020-12-15 *Translation:* (pakistan OR pak) (war OR firing OR army) (karachi OR police) since:2020-09-01 until:2020-12-15
सुप्रीम कोर्ट ने आदेश दिया कि मुस्लिम पुरुष और हिन्दू महिला का अब विवाह संभव नहीं *Translation:* Supreme Court orders that marriage of Muslim man and Hindu woman is no longer possible.	सुप्रीम (मुस्लिम OR मुसलमान) (शादी OR ब्याह OR विवाह) since:2020-07-01 until:2020-10-31 *Translation:* Supreme (muslim OR musalmaan) (marriage OR wedding OR matrimony) since:2020-07-01 until:2020-10-31
وفاة الفنان كاظم الساهر *Translation:* Death of the artist Kadim Al Sahir	(يمت OR توفى OR مات OR الموت OR موت OR وفاة) كاظم الساهر since:2021-12-01 until:2021-02-01 *Translation:* (Kadim AND Al Sahir) (Death OR Die OR Died) since:2021-12-01 until:2022-02-01
الفيفا توافق رسميا على إقامة كأس العالم كل سنتين *Translation:* FIFA officially agrees to host the world cup every two years	سنتين) (المونديال OR كأس العالم) (وافقت OR موافقة OR توافق) (فيفا OR الفيفا) OR عامين) since:2021-12-01 until:2022-02-01 *Translation:* FIFA (agrees OR agreed) (world cup) (two years) since:2021-12-01 until :2022-02-01

Table 13: Example English and Hindi claims with corresponding search queries. Queries are manually constructed to cast a broad net, retrieving both relevant and irrelevant messages containing the keywords.

Claim: ***The COVID-19 Vaccine has magnets or will make your body magnetic***
⋆ **Irrelevant**: *@dbongino is right. you can't tell people to wear a mask if the vaccines work. its like trying to put a north end of a magnet and trying to connect it to a north end another magnet., it will never work. #foxandfriends*
⋆ **Supporting**: *a friends family member got the covid vaccine and now she can put a magnet up to the injection site and the magnet stays on her arm.*
↳ **Supporting (only in context)**: *@ThisIsTexasFF Nano probes / tech / dust.*
Refuting (only in context): *@Newsweek Why the hell would they even bother with a high quantity of metal in the injection? And the amount that would be required to hold a magnet in place would be ridiculous.*
↳ **Refuting (only in context)**: *@pentatonicScowl @Newsweek I imagine the people making the claims don't fully understand how magnets work*
↳ **Supporting (only in context)**: *@AuracleDMG @pentatonicScowl @Newsweek Laugh now, cry later..*
↳ ⋆ **Refuting**: *@cis_kale Your point being? Even if these RNA vaccines contained ferric nanoparticles, they would not be in high enough concentrations to be able to hold a magnet in place. I suspect that blood itself has a higher concentration of ferric particles than the vaccine described in this paper*
⋆ **Querying**: *There is a #covid19 vaccine magnet test circulating on Tiktok, Is it really a thing?!!*
↳ **Supporting (only in context)**: *@Thepurplelilac well, 4 friends out of 9 can stick magnets to their arms so yeah, it's a thing*
⋆ **Discussing**: *@heggzigu @htmdnl too early to make any presumptions on either side. the truth has a way of exposing itself given enough time. bring a magnet to your vaccination appointment, see how the vaccine reacts with the magnet, maybe even bring a metal detector as well. would that convince you?*
⋆ **Discussing**: *Fauci: No Concern About Number of People Testing Positive After COVID-19 Vaccine. Spike Protein Vax is magnet for coronavirus. Originally used as turbo booster mounted on virus but too flimsy. Now injected in target in advance of infection, death rate 4X.*

Table 14: An example claim and its corresponding tweets from the 5 stance categories (best view in color): **Irrelevant**, **Refuting**, **Supporting**, **Discussing**, and **Querying**. ⋆ symbol indicates the tweets we directly retrieved from our query keyword search method. Indented lines with ↳ are replies to parent tweets.

وفاة الفنان كاظم الساهر *Translation:* Death of the artist Kadim Al Sahir
وفاة بطل المصارعة الحرة الأميركي أندرتيكر، جراء إصابته بفايروس كورونا *Translation:* The American wrestling champion, The Undertaker, dies of the Corona virus
انتخاب دروغبا رئيسا لإتحاد كرة القدم في ساحل العاج *Translation:* Drogba elected as president of the Ivory Coast Football Association
صورة لإضاءة برج خليفة بألوان علم لبنان بعد تتويج المنتخب اللبناني بلقب البطولة العربية لكرة السلة *Translation:* A picture shows the lighting of the Burj Khalifa in the colors of the Lebanese flag, after the Lebanese team won the Arab Basketball Championship title
البنك الدولي يُعلن إفلاس مصر بسبب العجز عن سداد ديونها *Translation:* The World Bank declares Egypt bankrupt due to the inability to pay its debts
اعتماد اليوان الصيني بديلًا عن الدولار الأميركي في أسواق روسيا *Translation:* The Chinese yuan is being adopted as an alternative to the US dollar in Russia's markets
روسيا تُهدِّد بقطع كابلات شبكات الإنترنت وإعادة العالم إلى العصر الحجري *Translation:* Russia threatens to cut Internet cables and send the world back to the Stone Age
الرئيس التركي رجب طيب أردوغان: سنعيد مليون سوري لبلادهم *Translation:* Turkish President Recep Tayyip Erdogan: "We will send a million Syrians back to their country"
إقالة المدير الفني لمنتخب المغرب وحيد خليلوزيتش *Translation:* The dismissal of Morocco's national team coach Vahid Halilhodzic
جو بايدن: الديانات الإبراهيمية الثلاث متشابهة ويمكن دمجها *Translation:* Joe Biden: The three Abrahamic religions are similar and can be combined
استُبعد منتخب إيران من مونديال قطر *Translation:* Iran has been excluded from the World Cup in Qatar
مجموعة إم بي سي توقف عقد الفنان رامز جلال بعد فشل برنامجه رامز موفي ستار *Translation:* MBC Group suspended the contract of the artist, Ramez Jalal, after the failure of his program "Ramez Movie Star"
مقتل الحكم الغامبي بكاري جاساما عقب تحكيمه مباراة الجزائر والكاميرون الأخيرة *Translation:* Gambian referee Bakary Gassama was killed after refereeing the last Algeria-Cameroon match
احتواء لقاح فايزر على أكسيد الجرافين القاتل ومؤامرة التطعيم للسيطرة على أعداد البشر وعدة ادعاءات أخرى *Translation:* Pfizer's vaccine contains deadly graphene oxide and the vaccination plot is to control human populations and several other allegations
محكة العدل الدولية بلاهاي تحكم بالغاء جميع اشكال التلقيح وتصنيعها وبيعها والغاء البروتوكول الصحي لمنظمة الصحة العالمية وتضع عدة شخصيات تحت الملاحقة القانونية الدولية منهم المدير العام لشركة فايزر بتهمة الابادة الجماعية *Translation:* The International Court of Justice in The Hague ruled to cancel all forms of vaccination, manufacture and sale and cancel the health protocol of the World Health Organization and put several personalities under international legal prosecution, including the Director General of Pfizer on charges of genocide.
تصريح الشاب خالد لقنوات بي إن سبورت: أتمنى فوز فريق عربي بكأس العرب *Translation:* Cheb Khaled's statement to beIN Sports: I hope an Arab team wins the Arab Cup
صورة للاعب كريستيانو رونالدو وهو يحمل كتاب أردوغان الجديد *Translation:* A picture of Cristiano Ronaldo shows him holding Erdogan's new book
صورة وحيد القرن بجانب رجل يزعم ناشروها أنها تظهر موت آخر حيوان وحيد القرن الأبيض الشمالي في العالم *Translation:* A picture of a rhinoceros next to a man whose publishers claim it shows the death of the last northern white rhino in the world
الفيفا توافق رسميا على إقامة كأس العالم كل سنتين *Translation:* FIFA officially agrees to hold the World Cup every two years
عشرات الآلاف من النساء في جميع أنحاء العالم أبلغن عن حدوث تغيرات في الدورة الشهرية بعد تلقي لقاح كورونا، مما أثار علامات استفهام بين الأطباء والعلماء *Translation:* Tens of thousands of women around the world have reported changes in their menstrual cycle after receiving the coronavirus vaccine, which has raised question marks among doctors and scientists
نهاية كورونا ومتغيراته ستكون عبر استخدام سم العقارب *Translation:* The end of covid and its variants will be through the use of scorpion venom
من يصاب بنزلات البرد لن يصيبه كورونا، كون الأجسام المضادة التي يكونها خلال إصابته بنزلات البرد تحميه داخليًا *Translation:* Whoever gets colds will not get covid, because the antibodies that he creates during colds protect him internally

Figure 2: Arabic Claims

महाराष्ट्र के अमरावती में हाल ही में हिंसा की घटना हुई थी. ऑप इंडिया ने एक तस्वीर शेयर करते हुए इसे मुसलमानों द्वारा हिन्दुओं की संपति जलाये जाने का दृश्य बताया. ये तस्वीर आधी थी. असल में ये दुकान एक मुस्लिम शख़्स की थी.	*Translation:* Recently there was an incident of violence in Amravati, Maharashtra. While sharing a picture, Op India described it as a scene of burning of Hindu property by Muslims. This was only a half of the full-picture. Actually this shop belonged to a Muslim man.
UP कांग्रेस सहित कई लोगों ने एक वीडियो शेयर करते हुए दावा किया कि योगी आदित्यनाथ ने केशव प्रसाद मौर्य को डांटा.	*Translation:* Many people, including the UP Congress, shared a video claiming that Yogi Adityanath scolded Keshav Prasad Maurya.
कुछ तस्वीरें और वीडियो शेयर करते हुए दावा किया जा रहा है कि तमिलनाडु के एक मंदिर में नियुक्त ईसाई पुजारी ने मंदिर में शराब पी कर मूर्तियों का अपमान किया.	*Translation:* Sharing some pictures and videos, it is being claimed that a Christian priest appointed in a temple in Tamil Nadu insulted the idols by drinking alcohol in the temple.
श्रीनगर मेडिकल कॉलेज की 100 छात्राओं की डिग्री रद्द कर दी गयी क्यूंकि उन्होंने 'पाकिस्तान ज़िंदाबाद' का नारा लगाया.	*Translation:* Degrees of 100 girl students of Srinagar Medical College were canceled because they raised the slogan 'Pakistan Zindabad'.
UP के बागपत में एक नाबालिग़ लड़की से बलात्कार की घटना हुई. इसके बाद सोशल मीडिया यूज़र्स ये दावा करने लगे कि 'जिहादियों' ने लड़की से रेप किया था.	*Translation:* A minor girl was raped in Baghpat, UP. After this, social media users started claiming that 'jihadis' had raped the girl.
UP के बदायूं में पेट्रोल पंप कर्मचारी का अपहरण कर लिया गया क्यूंकि उसके पास पर्याप्त कैश था.	*Translation:* A petrol pump employee was kidnapped in Badaun, UP because he had enough cash.
शाहरुख़ खान की एक तस्वीर शेयर की जा रही है जिसपर "वोट फॉर MIM " लिखा है. कहा जा रहा है कि उन्होंने असदुद्दीन ओवैसी की पार्टी के लिए वोट मांगा.	*Translation:* A picture of Shahrukh Khan is being shared on which "Vote for MIM" is written. It is being said that he sought votes for Asaduddin Owaisi's party.
आजमगढ़ में दुर्गा पूजा पंडाल हटाने की धमकी देने वाले शख़्स 'आदिल' को पुलिस ने गिरफ़्तार किया.	*Translation:* Police arrested a man 'Adil' who threatened to remove Durga Puja pandal in Azamgarh.

भारत को 'हिन्दू राष्ट्र' घोषित करने के लिए अयोध्या में भारी संख्या में राम भक्त जमा हो रहे हैं.	*Translation:* A large number of Ram devotees are gathering in Ayodhya to declare India as a 'Hindu Rashtra'.
4 साल की बच्ची से रेप के आरोपी का वीडियो शेयर किया जा रहा है. दावा किया जा रहा है कि उसका नाम अकरम खान है और वो AAP विधायक दिनेश मोहनिया का सोशल मीडिया कोऑर्डिनेटर है.	*Translation:* The video of a man accused of raping a 4-year-old girl is being shared. It is being claimed that his name is Akram Khan and he is the social media coordinator of AAP MLA Dinesh Mohaniya
12 सितम्बर को 26 रोहिंग्या मुसलमान फ़र्ज़ी आधार कार्डों के साथ पकड़े गए थे	*Translation:* 26 Rohingya Muslims were caught with fake Aadhar cards on 12 September
असम कांग्रेस के नेता 'अमजद अली' को हथियारों और गोलियों के साथ गिरफ़्तार किया गया.	*Translation:* Assam Congress leader 'Amjad Ali' was arrested with arms and bullets.
लखनऊ में एक तांगे पर पाकिस्तान का झंडा पेंट किया हुआ था. दैनिक जागरण ने कहा कि तांगेवाले ने 'पाकिस्तान ज़िंदाबाद' के नारे लगाए.	*Translation:* The flag of Pakistan was painted on a tonga in Lucknow. Dainik Jagran said that the Tangewale raised slogans of 'Pakistan Zindabad'.
हरियाणा में बजरंग दल के कार्यकर्ता एक मुस्लिम युवक को पीट रहे थे.	*Translation:* In Haryana, Bajrang Dal workers were beating a Muslim youth.
एक ठेलेवाला जग में पेशाब कर बर्तन में डालते हुए दिखता है. इसके साथ दावा किया जा रहा है कि ये शख़्स मुस्लिम समुदाय से है.	*Translation:* A street vendor is seen pouring urine in the jug and pouring it into the pot. With this it is being claimed that this person is from the Muslim community.
भारत पहली बार संयुक्त राष्ट्र सुरक्षा परिषद की अध्यक्षता करेगा.	*Translation:* India will chair the United Nations Security Council for the first time.
एमनेस्टी इंटरनेशनल पेगासस प्रोजेक्ट पर अपनी पहली रिपोर्ट से पीछे हट गया है.	*Translation:* Amnesty International has withdrawn from its first report on the Pegasus project.
अजमल खान' ने एक हिन्दू लड़की के पूरे परिवार को मार दिया क्यूंकि लड़की के घरवाले शादी के लिए राज़ी नहीं थे.	*Translation:* Ajmal Khan' killed the entire family of a Hindu girl because the girl's family members were not ready for the marriage
दिलीप कुमार की मौत के बाद सोशल मीडिया यूज़र्स ये दावा कर रहे उन्होंने अपनी संपति मुस्लिम वक्फ़ बोर्ड को दान दे दी	*Translation:* After the death of Dilip Kumar, social media users are claiming that he donated his property to the Muslim Waqf Board.
इतिहास में पहली बार 'मक्का-मदीना' में एक शिवलिंग देखा गया.	*Translation:* For the first time in history, a Shivling was seen in 'Mecca-Medina'.
तिरुपति बालाजी मंदिर के पुजारी के घर पड़ा छापा, 128 किलो सोना मिला	*Translation:* Raid at Tirupati Balaji temple priest's house, 128 kg gold found
संसद से कृषि विधेयकों के पास होते ही अडानी एग्री लॉजिस्टिक्स लिमिटेड पंजाब में अनाज गोदाम की स्थापना कर रही है.	*Translation:* Adani Agri Logistics Limited is setting up a grain warehouse in Punjab as soon as the Agriculture Bills are passed by the Parliament.
प्रदर्शनकारी बैनर पकड़े खड़े हैं जिसपर लिखा है "हमें कश्मीर नहीं चाहिए, विराट कोहली दे दो"	*Translation:* Protesters are standing holding banners that read "We don't want Kashmir, give us Virat Kohli".
सारण और महाराजगंज लोकसभा क्षेत्र में BDO की मौजूदगी में EVM से भरी एक गाड़ी स्ट्रांग रूम में घुसने की कोशिश कर रही थी.	*Translation:* A vehicle full of EVMs was trying to enter the strong room in the presence of BDO in Saran and Maharajganj Lok Sabha constituencies.
मुजफ्फरपुर रेलवे स्टेशन पर मृत मां के साथ खेलते जिस बच्चे का वीडियो जारी हुआ था, उस बच्चे के साथ शाहरुख की तस्वीर.	*Translation:* Shahrukh's picture with the child whose video was released playing with the dead mother at the Muzaffarpur railway station.
हाल में हुई एक घटना में सूडान में मुसलमानों ने फ्रांस के दूतावास में आग लगा दी.	*Translation:* In a recent incident, Muslims in Sudan set fire to the French embassy.
फ्रांस से नौकरी का प्रस्ताव पाने वाले ड्रोन वैज्ञानिक प्रताप एन एम को प्रधानमंत्री मोदी ने डीआरडीओ संस्था में नियुक्त किया है.	*Translation:* Drone scientist Pratap NM, who got a job offer from France, has been appointed by Prime Minister Modi in the DRDO organization.
स्वदेशी सामान के इस्तेमाल पर जोर देने वाले बाबा रामदेव ने खुद अपने घुटने का ऑपरेशन जर्मनी में करवाया है.	*Translation:* Baba Ramdev, who insisted on the use of indigenous goods, himself got his knee operated in Germany.
एयर इंडिया के विमान में सोशल डिस्टेंसिंग का ख्याल न रखने के लिए अटेंडेंट से बहस करते यात्रियों का वीडियो.	*Translation:* Video of passengers arguing with attendants for not taking care of social distancing in Air India plane.
जेएनयू में 5 जनवरी को हुई हिंसा के बाद एसएफआई कार्यकर्ता सूरी कृष्णन ने अपने घायल होने का दिखावा किया.	*Translation:* SFI activist Suri Krishnan pretends to be injured after the January 5 violence in JNU.
लखनऊ आत्मदाह की कोशिश के मामले में आसिफ नाम का शख़्स गिरफ्तार, युवक कांग्रेस नेता का पुत्र	*Translation:* A person named Asif arrested in Lucknow self-immolation case, Youth Congress leader's son
स्वस्थ आदमी को कोरोना पेशेंट बताकर अब तक 125 लोगों की किडनी निकालकर हत्या करने वाला डॉ देवेंद्र शर्मा गिरफ्तार।	*Translation:* Dr Devendra Sharma, who killed 125 people by removing their kidneys so far by calling a healthy person a corona patient, was arrested.
मैक्सिकन सांसद एंटोनियो गार्सिया ने संसद में अपने कपड़े उतार दिए	*Translation:* Mexican MP Antonio García took off his clothes in parliament
पाकिस्तान में ग्रह युद्ध छिड़ गया है। पाक आर्मी और कराची पुलिस के बीच जबरदस्त फायरिंग शुरू हो गई है।	*Translation:* In-house war has broken out in Pakistan. Heavy firing has started between Pak Army and Karachi Police.
पाकिस्तान के पेशावर के मदरसे में बम बनाया जा रहा था।	*Translation:* bomb was being made in a madrassa in Peshawar, Pakistan.
गुजरात के सूरत में लोग सड़क पर राशन दो या गोली मार दो के नारे लगा रहे हैं।	*Translation:* In Surat, Gujarat, people are raising slogans of 'give ration or shoot' on the road.
चीन ने कोर्ट से कोरोना संक्रमित 20,000 मरीज़ों को मारने की मंजूरी मांगी।	*Translation:* China asked the court for approval to kill 20,000 corona-infected patients.
सुप्रीम कोर्ट ने आदेश दिया कि मुस्लिम पुरुष और हिन्दू महिला का अब विवाह संभव नहीं	*Translation:* Supreme Court orders that marriage of Muslim man and Hindu woman is no longer possible
अर्जुन मिश्रा नामक व्यक्ति ने साइकिल गर्ल ज्योति की हत्या कर दी।	*Translation:* A person named Arjun Mishra killed cycle girl Jyoti.

Table 15: Hindi Claims

2020 is a year of global cooling, or we are entering into a period of global cooling
Ancient marine fossils on top of Mount Everest are proof of a Great Flood in the past
Astrazeneca means, in latin, “kill the stars” or “weapon that kills”
Asymptomatic transmission is a concept invented in 2020 for the COVID-19 pandemic
Bill Gates vaccine in India paralyzed hundreds of thousands of people in India
Bill Gates’ Zika vaccine for Brazil was actually the cause of microcephaly and Zika to occur in Brazil
COVID-19 nasal swabs harm the brain or penetrate the blood brain barrier
Drinking Cold Water is unhealthy, causes cancer, and causes heart attacks
Due to the cold winter in Germany, there was so much snow that the solar panels and wind turbines of Germany were rendered COMPLETELY useless
Easter is a celebration for the Mediterranean Goddess Ishtar
Ethylene Oxide, a known carcinogen, is in COVID-19 test swabs and in face masks
Glaxosmithkline is linked to the Wuhan lab that “released” COVID-19
Jet Fuel is a hoax, most jets run off of compressed air
Life insurance is voided, and you will not get life insurance if you take the COVID-19 vaccine
The Red Cross is withholding bushfire money from Australia
The Three gorges dam in China flooded or leaked in 2020 or 2021, killing thousands
The false positive rate for a COVID-19 test is very high
There has been no net warming of the planet in the past 22 years
Using soap and a cotton ball is the best way to remove a tick
Western Australia and Australia has admitted that the COVID19 vaccine is poison
50,000 Chinese soldiers were killed trying to invade Maine
California’s new bill decriminalizes pedophilia
Canada has internment camps for COVID-19
Canada is planning military intervention if Trump does not leave office
Canada’s Supreme Court is hearing a case about COVID-19 crimes, crimes against humanity, and genocide
Canada’s elections are suspended indefinitely
Holy communion is banned in Toronto
Justin Trudeau advocates for beastiality, or beastiality is legal in Canada
Justin Trudeau and the UN have revealed plans for the “Great Reset”, which will ruin society
Justin Trudeau supports pedophilia
Justin Trudeau went to barbados during Christmas, breaking COVID-19 guidelines
Masks cause cancer
Most ANTIFA members are school teachers
Polysorbate 80 and Potassium Chloride are in the COVID-19 vaccine
Quebec is exempt from the Federal carbon tax and pays less for carbon tax
The COVID-19 Vaccine causes cancer
The COVID-19 vaccine is radioactive or has radioactive substances
The Canadian Election is full of fraud because Justin Trudeau used Dominion Voting Systems
The husband (Richard Choi) of Toronto’s COVID response chief (Eileen de Villa) has been arrested for fraud
These COVID-19 testing trucks have the logo Anubis, the Egyptian God of Death
A childs presence at school is implied consent for administering the COVID vaccine
Bees use acoustic levitation to fly
Boris Johnson’s father Stanley Johnson wrote a book called The Virus in the 1980s that predicted the COVID-19 vaccine
COVID-19 does not spread on surfaces
Chemicals, chips, and plastic are in masks
Drinking water from a copper cup has health benefits
Facebook supports pedophilia, or Facebook has an ad advocating for pedophilia
Fossil fuels are actually a mineral, and Rockefeller coined the term Fossil Fuels to introduce the idea of “scarcity”
Hunza people are free from cancer
Jacinda Ardern and the New Zealand Labour party bribed the New Zealand media to win the reelection and there is election fraud in New Zealand
Low Flu cases this year prove that the pandemic is fake
NASA added a 13th zodiac sign Ophichus
New Zealand Labour Party is talking about a forced coronavirus vaccination
New Zealand revoked the human right to grow food
Plant seeds can restructure their DNA and become a “super food” if you spit on them
Swine flu is more dangerous than COVID-19
The Bacterial meningitis vaccine caused the outbreak of the “Spanish Flu”
The Pfizer COVID-19 vaccine has a new upgrade with a Microsoft chip
The etymology of Hangover derives from the historical practice of paying 2 cents to sleep on a rope, hanging there
The spike protein in the COVID-19 vaccine is dangerous
As of May 2021 Singapore is in a circuit breaker lockdown
Black Fungus is extremely contagious and is a mutant and variant of COVID-19
COVID vaccine causes a brain hemorrhage as seen from a doctor
COVID-19 vaccine causes seizures
COVID-19 vaccine creates these new coronavirus variants
COVID-19 vaccine passports will be mandatory for travel
Countries are using military helicopters to spray pesticides to disinfect against COVID-19
Finland is promoting a 4 day work week
Infrared thermometers hurt the pineal gland
Jack Ma is missing
Killed-virus vaccines are superior to mRNA and have always been perfectly safe
Prion in COVID-19 vaccine causes neurodegenerative diseases like Alzheimers
Singapore followed Trump’s COVID-19 Guidelines
The COVID-19 vaccine causes Bells Palsy as a side effect
The COVID-19 vaccine increases the chance of heart attack
The TraceTogether App uses GPS data to track location
The app Muslim Pro is selling user data to the US military
The elderly are dying of the COVID-19 vaccine
There is a new COVID-19 variant in Singapore that is extremely dangerous
Trump wrote “Joe, you know I won” in a letter to President Biden
5g testing is the cause for the second wave of COVID-19 cases in India
Amul ice cream, Amul products, and E471 is made from pig fat but is branded as vegetarian
An empty Nebuliser can be used as a substitute for oxygen cylinders
Aspidosperma-Q is a substitute to medical oxygen
Bharat Biotech’s Covaxin has been approved for usage for children above 12 years old
Bitter melon, bitter gourd, or bitter melon juice is a cure for COVID-19
Burning Ghee will produce Oxygen
COVID-19 is able to be transmitted by houseflies
Clapping will kill COVID-19 due to noise and vibrations
Dead birds are evidence of 5g trials in India
Drinking coconut water kills cancer cells and cures cancer
Hospital staff are intentionally killing patients in India
Lemon drops in the nose can kill COVID-19
Luc montagnier claimed that vaccinated people will die in 2 years
Nostradamus predicted the outbreak of COVID-19 hundreds of years ago
Raw onions with rock salt will cure COVID-19
Sonu Sood promises to support Hamas and Palestine
The Polio vaccine was introduced in India in 1995 or later
The majority of the education ministers of India are muslim; the first 5 education ministers of india are muslim
World Bank exported COVID-19 tests in 2018 or 2017 before COVID-19 pandemic occurred
AZ Vaccine is banned in multiple countries
COVID was planned
COVID-19 vaccine is an experimental vaccine
Covid vaccine cause blood clot
Cucumbers cures cancer and also prevents cancer
Dandelion Root and Dandelion Extract cures cancer
Formaldehyde is in the COVID vaccine and it is enough to cause cancer
Japan refuses to receive blood from anyone who has been vaccinated
Mandatory vaccine for care workers
Morgellon’s is a real physical disease and it is caused by nanotechnology, clothing fibers, fibers in the mask, the COVID-19 vaccine, or COVID-19 PCR tests

Priti Patel spent £77,000 of the Home Office's money on eyebrows and beauty products
Rapper DMX had the COVID-19 vaccine before his death (and causing his death)
Soy sauce imports from Japan will be cheaper next year because of Brexit
The COVID vaccine kept killing all the animals it was tested on
The COVID-19 vaccine contains Chloroform, AKA SM-102
The COVID-19 vaccine has magnets or makes you magnetic
The COVID-19 vaccine violates the Nuremberg codes
The Genocide Act of 1969 was repealed in May 2021, so genocide is now legal
The UK is doing the worst for dealing with COVID-19 because they have the highest COVID-19 death toll in all of Europe
The suicide rate increased during COVID-19 lockdown
Acid attack victims have been included in the list of people considered as disabled.
Afghanistan has more hectares of opium poppies planted today than they had before we started this war.
African-Americans don't use drugs at a higher level than whites but wind up going to prison six times more.
All of our taxpayers are paying roughly 22% of the U.N. budget.
Asia has less freshwater per capita than any other continent, except Antarctica.
British voters "under 50, especially millennials, overwhelmingly voted to stay," in the European Union. "It was older voters who voted to leave."
Cervical cancer in Africa is "fully preventable with basic education, screening and vaccines."
Forty-seven percent of Americans pay no income tax.
Forty-three million Americans are on food stamps.
Global food demand is expected to increase by 50 to 97 percent by 2050.
In Africa, a child dies every minute because of (malaria).
In Malaysia many of the workers are indentured servants because their passports are taken away when they come into this country and are working in slave-like conditions.
In South Sudan, more teenage girls die in childbirth than finish high school.
Malawi has just 300 doctors for 16 million people.
More people die from indoor air pollution than from malaria, HIV/AIDS and TB combined.
Research even shows that sending more girls to school can boost an entire country's GDP.
Studies have consistently failed to establish the existence of a link between the harshness of a country's drug laws and its levels of drug use.
Swaziland has the highest rate of HIV infection in the world.
The world's 62 richest people own the same wealth as the 3.6 billion poorest.
Today, 27 million people are enslaved.
Bathing in sea water and injecting sea water is a cure to COVID-19
COVID-19 dies at various low temperatures
COVID-19 is linked to a substance called "adrenochrome"
Coronil is a cure to COVID-19
Fetal cells are in the COVID-19 vaccine
George Soros owns the Wuhan lab that released COVID-19 to the world
Hydroxychloroquine is approved and safe to treat COVID-19
Inhaling Camphor and Ajwain helps with increasing Oxygen levels in the body
Ivermectin is safe to be approved to treat COVID-19
Madagascar found a cure to COVID-19 and will be distributing it to other countries
Mosquitoes transmit COVID-19
Mustard and Mustard Oil cures COVID-19
Smoking Nicotine or Marijuana protects and cures you from COVID-19
Sweden stops using PCR tests to diagnose COVID-19
The Simpsons predicted COVID-19
The pandemic is fake, health officials and hospitals are administering empty syringes as the "COVID-19" vaccine
Turmeric is a treatment or cure to COVID-19
Vaccines cause miscarriage
Vaccines will make you sterile
Wearing a mask causes pleurisy in an individual
Avon employees will relocate victims of Domestic Violence
CNN said that the Boulder, CO Shooter was Morally White
Church bells rang in Paris to celebrate Bidens election win
Dominion Voting Systems is manipulating the election and deleting Trump votes
Dr. Deborah Birx has been arrested by the military for spreading COVID-19 conspiracies
George Soros owns Parler
Hillary Clinton is trafficking children aboard the Ever Green Ship
Hillary Clinton was convicted of murder by a military tribunal
Hunter Biden is teaching a "Fake News" class
Jacques Attali Encourages Pandemic Driven Euthanasia in 1981
Joe Biden is a wanted felon in Ukraine
Joe Biden said that white republican men are a bigger danger than ISIS and Al-Qaeda
Joni Ernst had a liberal fantasy island poster on the senate floor that included messages like "sex blimps" and "abolish lasagna"
Lin Wood found Trump at the White House and Biden missing
Luciferase is the name or ingredient in the COVID-19 vaccine
Mail-in Ballots contain Quantum Blockchain Watermarks
Maricopa County Elections deleted their database
Mary Tyler Moore Died in 2021
Masks cause bacterial pneumonia
Paris Hilton wore a "Stop Being Poor" shirt
The CDC is a private, non-profit organization
The Grapes of Wrath is translated as the Angry Raisins in several languages
There are real tickets sold online for Trump's 2nd inauguration in August 2021
There are thousands of Georgia Ballots that are ineligible
There is a Hammer and Scorecard operation to alter the Election results
There is a code and post that allows you to circumvent the facebook algorithm and news posts to see your friend's posts
There were no mass shootings under Trump
Trump will be reinstated as US president in August 2021
Trump wore his pants backwards at a rally
UNICEF reported that blocking kids from porn violates human rights

Table 16: English Claims