PANACEA: An Automated Misinformation Detection System on COVID-19

Runcong Zhao[1,3], **Miguel Arana-Catania**[5], **Lixing Zhu**[1,3], **Elena Kochkina**[2,4], **Lin Gui**[3], **Arkaitz Zubiaga**[2], **Rob Procter**[1,4], **Maria Liakata**[1,2,4], **Yulan He**[1,3,4]
[1]University of Warwick, [2]Queen Mary University of London
[3]King's College London, [4]The Alan Turing Institute, [5]Cranfield University
{runcong.zhao, yulan.he}@kcl.ac.uk

Abstract

In this demo, we introduce a web-based misinformation detection system PANACEA on COVID-19 related claims, which has two modules, *fact-checking* and *rumour detection*. Our *fact-checking* module, which is supported by novel natural language inference methods with a self-attention network, outperforms state-of-the-art approaches. It is also able to give automated veracity assessment and ranked supporting evidence with the stance towards the claim to be checked. In addition, PANACEA adapts the bi-directional graph convolutional networks model, which is able to detect rumours based on comment networks of related tweets, instead of relying on the knowledge base. This *rumour detection* module assists by warning the users in the early stages when a knowledge base may not be available.

1 Introduction

The dangers of misinformation have become even more apparent to the general public during the COVID-19 pandemic. Following *false* treatment information has led to a high number of deaths and hospitalisations (Islam et al., 2020). Manual verification can not scale to the amount of misinformation being spread, therefore there is a need to develop automated tools to assist in this process.

In this work, we focus on automating misinformation detection using information from credible sources as well as social media. We produce a web-based tool that can be used by the general public to inspect relevant information about the claims that they want to check, see supporting or refuting evidence, and social media propagation patterns.

For *false* information, the commonly used and relatively reliable method for automated veracity assessment is to check the claim against a verified knowledge base, which we call *fact-checking*. Previous works such as EVIDENCEMINER (Wang et al., 2020b), PubMed[1] and COVID-19 fact-checking sites recommended by the NHS[2] are all designed to retrieve related documents/sentences from a reliable knowledge base. However, this approach leaves users to summarise a large amount of potentially conflicting evidence themselves. PANACEA, which is supported by novel natural language inference methods (Arana-Catania et al., 2022), is instead able to provide automated veracity assessment and supporting evidence for the input claim. In addition, previous works retrieve results using entities in the input claim, and thus often include results related to a keyword in the input claim instead of the whole query, while PANACEA considers the whole query for better result. The supporting pieces of evidence are also ranked by their relevance score and classified according to their stance towards the input claim.

In addition to *false* information, truthful information can also be misused to harm competitors or gain attention on social media (Pennycook et al., 2020; Tsfati et al., 2020). However, the latter is harder to be found by checking reliable knowledge bases as those are focused on *false* information. Regarding this issue, previous work has analysed the spread of misinformation using features such as stance (Zhu et al., 2021), sentiment, topics, geographical spread, the reliability of external links included in the tweet (Sharma et al., 2020), origin and propagation networks (Finn et al., 2014). However, it is still hard for users to identify rumours by directly looking at those features. Previous research shows that the propagation pattern is different between fake and real news, which would offer additional features for early detection of misinformation on social media (Zhao et al., 2020). PANACEA extends this by using tweets' propagation patterns to identify rumours. *Rumour detection* is not as reliable as *fact-checking*, but it generalises the system to various situations that *fact-checking*

[1] https://www.ncbi.nlm.nih.gov/pmc/

[2] https://library.hee.nhs.uk/covid-19/coronavirus-%28covid-19%29-misinformation

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics System Demonstrations, pages 67–74
May 2-4, 2023 ©2023 Association for Computational Linguistics

cannot cover: First, *true* or *unverified* information with *intent to harm*; Second, scenarios where no verified knowledge database is available. *Rumour detection* cannot prove the truth of a claim but may alert the user about claims with a high risk of being misinformation.

Previous work have either retrieved tweets from a short fixed time period (Sharma et al., 2020) or search recent tweets (Finn et al., 2014), which is limited by Twitter to only the last 7 days. We instead maintain an updated database which is constituted of an annotated tweets dataset with popular claims and an unlabelled streaming of COVID-19 related tweets that are crawled and selected periodically to update the dataset. Besides building on the various analytic functionalities used in previous work, PANACEA improves the architecture of these elements and adds extra features to the updated dataset for more efficient results.

A screencast video introducing the system[3], illustrating its use in the checking of a COVID-19 claim, and the demo[4] are also available online. The system can be easily adapted to other claim topics.

PANACEA covers various types of misinformation detection related to COVID-19 with the following contributions:

- We built a new web-based system, PANACEA, which is able to perform both *fact-checking* and *rumour detection* with natural language claims submitted by users. The system includes visualisations of various statistical analyses of the results for a better user understanding.

- PANACEA performs automated veracity assessment and provides supporting evidence that can be ranked by various criteria, supported by novel natural language inference methods. The system is able to manage multiple user requests with low latency thanks to our development of a queuing system.

- PANACEA is able to perform automated rumour detection by exploiting state-of-the-art research on propagation patterns. The system uses an annotated dataset and streams of COVID-19 tweets are collected to maintain an updated database.

2 Datasets

The following datasets are used in the project:

Knowledge Database This is used for fact-checking, and includes COVID-19 related documents from selected reliable sources [5]. The documents were cleaned and split into 300 token paragraphs to construct a reliable knowledge database, whose supporting documents are retrieved and visualised in our system.

PANACEA Dataset (Arana-Catania et al., 2022), constructed from COVID-19 related data sources[6] and using BM25 and MonoT5 (Nogueira et al., 2020) to remove duplicate claims. This dataset includes 5,143 labelled claims (1,810 *False* and 3,333 *True*), and their respective text, source and claim sub-type.

COVID-RV dataset In order to fine-tune our model, we constructed a new COVID-19 related propagation tree dataset for rumour detection. Similar previous datasets are Twitter15 and Twitter16 (Ma et al., 2018), which are widespread tweets' propagation trees with rumour labels, however, they are not COVID-19 related. Our dataset has been constructed by extending COVID-RV (Kochkina et al., 2023), including *the number of retweets*, *user id*, *post time*, *text*, *location* and *tweet reply ids* as metadata for each tweet. Each tree is annotated with a related claim chosen from our claim dataset and a stance label (chosen from *Support* or *Refute*) towards its related claim. Such a stance label for each tree is purely based on the content of the source tweet. In COVID-RV the conversations are annotated as either *True* or *False* based on the veracity of the claim and the stance of the source tweet towards it. Tweets supporting a false claim or challenging a true claim are annotated as *False*, tweets supporting true claims or challenging a false claim are annotated as *True*. Twitter15 and Twitter16 datasets also contain *Unverified* conversations, which are discussing claim that are neither confirmed or denied.

COVID Twitter Propagation Tree (Live) Besides the last dataset constructed for fine-tuning,

[3] https://www.youtube.com/watch?v=D1PN8_9oYso

[4] https://panacea2020.github.io/

[5] *Centers for Disease Control and Prevention* (CDC), *European Centre for Disease Prevention and Control* (ECDC), *WebMD* and *World Health Organisation* (WHO)

[6] Corona VirusFacts Database, CoAID dataset (Cui and Lee, 2020), MM-COVID (Li et al., 2020), CovidLies (Hossain et al., 2020), TREC Health Misinformation track and TREC COVID challenge (Voorhees et al., 2021)

PANACEA also runs a crawler to collect a stream of COVID-19 tweets that are used to maintain an updated database. This live dataset is not annotated, instead, it is labelled by the pre-trained rumour detection model. As the Twitter's search API does not allow retrieval of tweets beyond a week window, we retrieve COVID-19 related historical tweets based on the widely used dataset of COVID-19-TweetIDs (Chen et al., 2020), which contains more than 1 billion tweet IDs. Considering the size of the dataset, and for the storage and retrieval efficiency, we filtered out the less popular tweets with limited impact. To date, more than 12k propagation trees have been collected, starting from January 2020. For each tweet, its pseudo rumour label is generated by the trained model.

3 Architecture of PANACEA

Figure 1 shows an overview of PANACEA, including two functions: *fact-checking* and *rumour detection* for COVID-19. For *fact-checking*, there are three modules: (1) resource allocation system; (2) veracity assessment; and (3) supporting evidence retrieval. PANACEA also supports a unique function, *rumour detection* by propagation patterns, which has the following modules: (1) tweet retrieval; (2) rumour detection; and (3) tweet meta-information analysis.

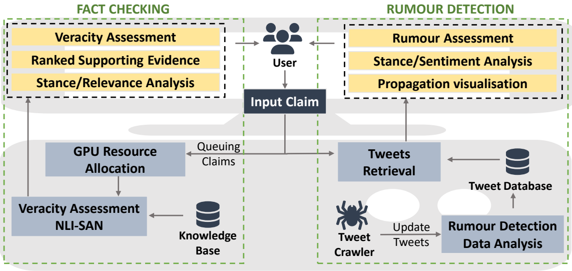


Figure 1: Architecture of PANACEA

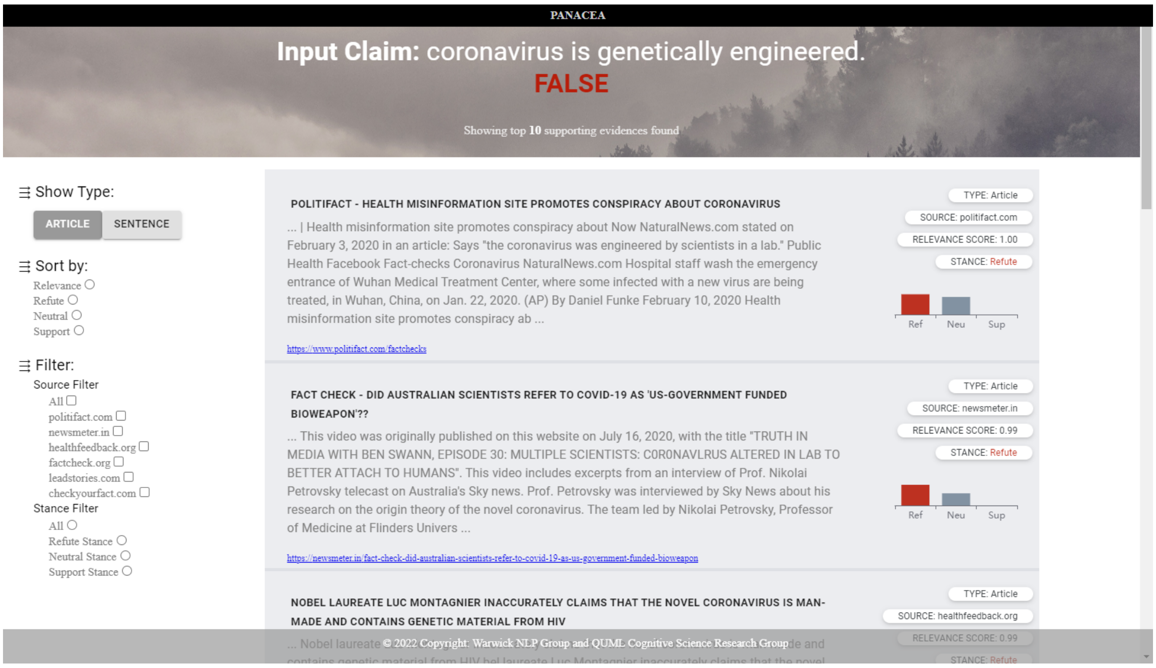
3.1 Fact-Checking

Resource Allocation System Users can input natural language claims into our system, and PANACEA provides autocompleted input guesses based on the current input and the claims dataset. Claim autocompletion can help users to input the claim faster and the results included within the claims dataset can be pre-computed for faster retrieval. However, if the user cannot find what they would like to check through the claims dataset, the new claim would be passed to our model for real-time evaluation. Veracity assessment and evidence retrieval are based on our natural language inference model NLI-SAN (Arana-Catania et al., 2022), which needs GPU resources to run. Therefore we built a queuing system that manages the resources and queues the claims while the GPUs are being used. The results are sent to the user. To avoid duplicate searches, a temporary copy of this result is saved in our database based on the user's IP address until the user searches for a new claim or the saved period expires.

Veracity Assessment PANACEA is supported by NLI-SAN (Arana-Catania et al., 2022), which incorporates natural language inference results of claim-evidence pairs into a self-attention network. The input claim c is paired with each retrieved relevant evidence e_i to form claim-evidence pairs, where the relevant evidences are the retrieved sentences as described in the following paragraph. Each claim-evidence pair (c, e_i) is fed into both a RoBERTa-large[7] model to get a representation S_i and into a RoBERTa-large-MNLI[7] model to get a probability triplet I_i of stance (*contradiction*, *neutrality*, or *entailment*) between the pair. Next, S_i is mapped to a Key K and a Value V, while I_i is mapped onto a Query Q. $(Q, K, V)_i$ forms the input of the self-attention layer and the outputs O_i for all the claim-evidence pairs are concatenated together. The output is then passed to a MLP layer to get the veracity assessment result (*True* or *False*) as shown in Figure 2.

Supporting Evidence Retrieval This module includes three parts: document retrieval, sentence retrieval and corresponding meta-data generation. Multi-stage retrieval is applied, retrieving first the top 100 relevant documents with BM25, that then are re-ranked by MonoT5 (Nogueira et al., 2020) and the top 10 documents are selected. For each of those documents, the top 3 sentences are selected. Both documents and sentences are ranked by their relevance score, which is the cosine similarity between the documents/sentences and the input claim embeddings. Each of those texts are represented through embeddings obtained using Sentence-Transformers with the pre-trained model MiniLM-L12-v2 (Wang et al., 2020a). The corresponding metadata of the supporting documents, including type, source, relevance score, and stance towards the claim are also shown, together with the ranked documents/sentences. Users can also

[7] https://huggingface.co/



PANACEA

Input Claim: coronavirus is genetically engineered.

FALSE

Showing top **10** supporting evidences found

☰ Show Type:

ARTICLE SENTENCE

☰ Sort by:

Relevance ○
Refute ○
Neutral ○
Support ○

☰ Filter:

Source Filter
All □
politifact.com □
newsmeter.in □
healthfeedback.org □
factcheck.org □
leadstories.com □
checkyourfact.com □
Stance Filter
All ○
Refute Stance ○
Neutral Stance ○
Support Stance ○

POLITIFACT - HEALTH MISINFORMATION SITE PROMOTES CONSPIRACY ABOUT CORONAVIRUS

... | Health misinformation site promotes conspiracy about Now NaturalNews.com stated on February 3, 2020 in an article: Says "the coronavirus was engineered by scientists in a lab." Public Health Facebook Fact-checks Coronavirus NaturalNews.com Hospital staff wash the emergency entrance of Wuhan Medical Treatment Center, where some infected with a new virus are being treated, in Wuhan, China, on Jan. 22, 2020. (AP) By Daniel Funke February 10, 2020 Health misinformation site promotes conspiracy ab ...

https://www.politifact.com/factchecks

TYPE: Article
SOURCE: politifact.com
RELEVANCE SCORE: 1.00
STANCE: Refute

FACT CHECK - DID AUSTRALIAN SCIENTISTS REFER TO COVID-19 AS 'US-GOVERNMENT FUNDED BIOWEAPON'??

... This video was originally published on this website on July 16, 2020, with the title "TRUTH IN MEDIA WITH BEN SWANN, EPISODE 30: MULTIPLE SCIENTISTS: C0R0NAVLRUS ALTERED IN LAB TO BETTER ATTACH TO HUMANS". This video includes excerpts from an interview of Prof. Nikolai Petrovsky telecast on Australia's Sky news. Prof. Petrovsky was interviewed by Sky News about his research on the origin theory of the novel coronavirus. The team led by Nikolai Petrovsky, Professor of Medicine at Flinders Univers ...

https://newsmeter.in/fact-check-did-australian-scientists-refer-to-covid-19-as-us-government-funded-bioweapon

TYPE: Article
SOURCE: newsmeter.in
RELEVANCE SCORE: 0.99
STANCE: Refute

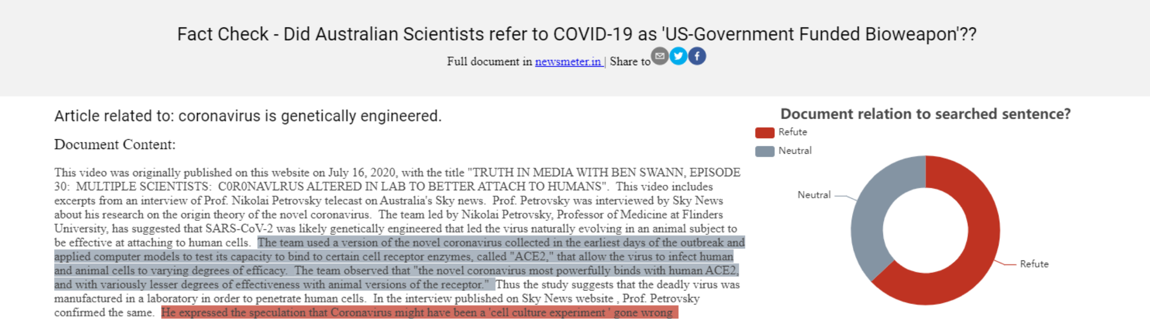
NOBEL LAUREATE LUC MONTAGNIER INACCURATELY CLAIMS THAT THE NOVEL CORONAVIRUS IS MAN-MADE AND CONTAINS GENETIC MATERIAL FROM HIV

... Nobel laureate de and contains genetic material from HIV bel laureate Luc Montagnier inaccurately claims that the novel

TYPE: Article
SOURCE: healthfeedback.org
RELEVANCE SCORE: 0.99
STANCE: Refute

© 2022 Copyright: Warwick NLP Group and QUML Cognitive Science Research Group

Figure 2: Fact checking result with input claim: *coronavirus is genetically engineered.*



Fact Check - Did Australian Scientists refer to COVID-19 as 'US-Government Funded Bioweapon'??

Full document in newsmeter.in | Share to

Article related to: coronavirus is genetically engineered.

Document Content:

This video was originally published on this website on July 16, 2020, with the title "TRUTH IN MEDIA WITH BEN SWANN, EPISODE 30: MULTIPLE SCIENTISTS: C0R0NAVLRUS ALTERED IN LAB TO BETTER ATTACH TO HUMANS". This video includes excerpts from an interview of Prof. Nikolai Petrovsky telecast on Australia's Sky news. Prof. Petrovsky was interviewed by Sky News about his research on the origin theory of the novel coronavirus. The team led by Nikolai Petrovsky, Professor of Medicine at Flinders University, has suggested that SARS-CoV-2 was likely genetically engineered that led the virus naturally evolving in an animal subject to be effective at attaching to human cells. The team used a version of the novel coronavirus collected in the earliest days of the outbreak and applied computer models to test its capacity to bind to certain cell receptor enzymes, called "ACE2," that allow the virus to infect human and animal cells to varying degrees of efficacy. The team observed that "the novel coronavirus most powerfully binds with human ACE2, and with variously lesser degrees of effectiveness with animal versions of the receptor." Thus the study suggests that the deadly virus was manufactured in a laboratory in order to penetrate human cells. In the interview published on Sky News website , Prof. Petrovsky confirmed the same. He expressed the speculation that Coronavirus might have been a 'cell culture experiment ' gone wrong

Figure 3: The detail page of user selected supporting document

filter or re-rank the result using the metadata. An example of documents retrieved is shown in Figure 2 and the corresponding detailed information visualisation is shown in Figure 3. On the details page, the whole document text is shown with the top 3 relevant sentences highlighted by their stance towards the input claim. The stance distribution, described in the veracity assessment module is also visualised.

3.2 Rumour Detection

Another approach to detecting rumours that has been found to be effective (Ma et al., 2018; Tian et al., 2022) is modelling user comments and propagation networks. Next we describe the relevant rumour detection modules of our system.

Claim-related tweets retrieval Similar to the fact-checking module, this module includes an autocomplete function for the user's natural language input claim that guesses the input from our claims dataset. The results for existing claims are also pre-computed to retrieve tweets faster. For a claim that is not in our claim dataset, we use BM25 to retrieve the related propagation trees from the large Twitter propagation tree database maintained by the active Twitter crawler.

Rumour Assessment and Data Analysis PANACEA adapts a bi-directional graph convolutional networks model (BiGCN) (Bian et al., 2020) to perform rumour detection, which is trained on Twitter16 and fine-tuned on our annotated propagation trees. The reason we chose BiGCN is that it behaves relatively better compared with

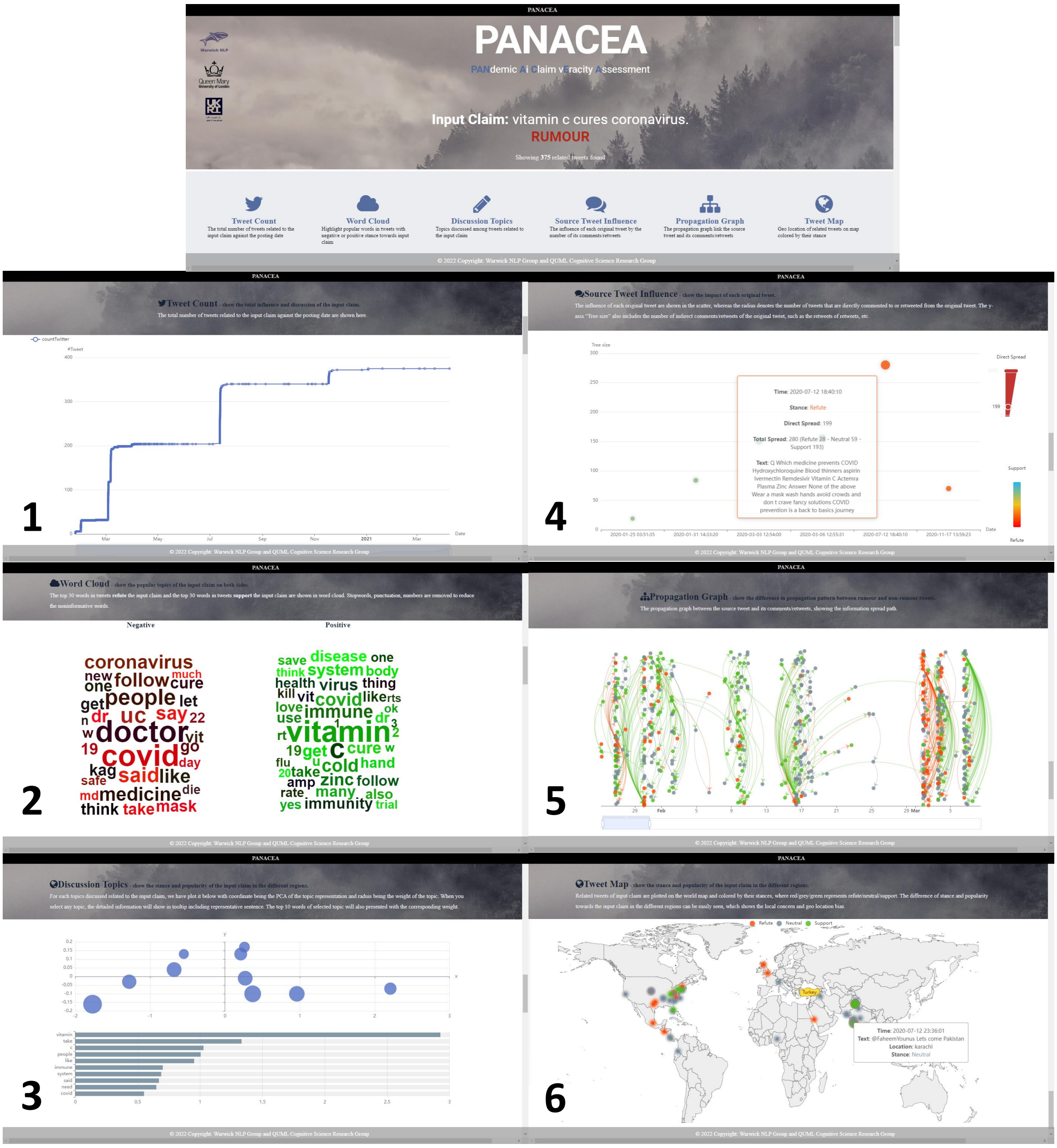

PANACEA
PANdemic Ai Claim veracity Assessment
Input Claim: vitamin c cures coronavirus.
RUMOUR
Showing 375 related tweets found
Tweet Count
Word Cloud
Discussion Topics
Source Tweet Influence
Propagation Graph
Tweet Map
1
2
3
4
5
6
Negative
Positive
Refute
Neutral
Support
Turkey
© 2022 Copyright: Warwick NLP Group and QUML Cognitive Science Research Group

Figure 4: Rumour detection result with input claim: *vitamin c cures coronavirus.*

other models in cross-dataset evaluation (Kochkina et al., 2023). For an input claim, the system gives the rumour detection result generated by the weighted average of propagation trees' rumour assessment label, $\frac{\sum_{i \in T} n_i r_i}{\sum_{j \in T} n_j}$, where T is the set of retrieved propagation trees. We generate the sentiment labels of each tweet by VADER[8] and stance of tweet towards the input claims by natural language inference (Nie et al., 2020).

Twitter propagation visualisation As shown in Figure 4, PANACEA has six modules, which use the metadata we crawled from the tweet and generated from data analysis to visualise the propagation pattern:

1. Tweet Count, showing the total number of tweets related to the input claim against the posting date, and aiming to reflect the total influence and scale of discussion of the claim.
2. Word Cloud, showing the top 30 words in tweets refuting the input claim and the top 30 words in tweets supporting the input claim.

[8] `https://www.nltk.org/api/nltk.sentiment.vader.html`

Stopwords, punctuation, and numbers are removed to reduce non-informative words.

3. Discussion Topics, building on Latent Dirichlet Allocation (LDA), where each topic is encoded by COVID-Twitter-BERT [9] and the representative tweet is selected by its embedding similarity with respect to the topic. Principal component analysis (PCA) is applied to visualise each topic. Top 10 words and corresponding weights of the chosen topic are shown in a bar chart.
4. Tweet Spread, showing the influence of each original tweet in the scatter plot, where the radius denotes the number of tweets that are direct comments or retweets from the original tweet. The y-axis "Total Spread" also includes the number of indirect comments/retweets of the original tweet, such as the retweets of retweets, etc.
5. Propagation Graph, showing the propagation graph between the source tweet and its comments, showing the information spread path. 5 other claims are randomly chosen from popular claims for users to compare propagation patterns. This module aims to visualise propagation graphs in a straightforward way and help users see the difference between trees of different types.
6. Tweet Map. Related tweets to the input claim are plotted on the world map and coloured by their stances, where *red/yellow/blue* represents *refute/neutral/support*. The difference in stance and popularity towards the input claim in the different regions can be easily seen, which shows the local context and geo-location bias.

4 Evaluation Results

Fact-Checking We investigate the performance of our system in document retrieval and veracity assessment in (Arana-Catania et al., 2022). Table 1 shows that combining BM25 and MonoT5 is the most effective approach for document retrieval of the selected techniques. In addition, Figure 5 shows that NLI-SAN achieves similar performance with KGAT (Liu et al., 2020), while having a simpler architecture for the application, and outperforms GEAR (Zhou et al., 2019).

	AP@5	AP@10	AP@20	AP@100
BM25	0.54	0.56	0.58	0.62
BM25+MonoBERT	0.52	0.55	0.58	0.62
BM25+MonoBERT	**0.55**	**0.58**	**0.60**	**0.62**
BM25+RM3+MonoT5	0.51	0.53	0.55	0.57

Table 1: Document retrieval on the PANACEA dataset.

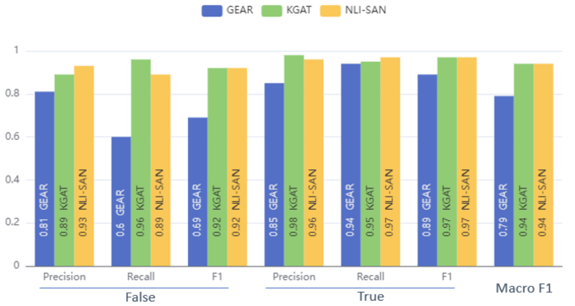


Figure 5: Veracity classification on the PANACEA dataset.

Rumour Detection As shown in Figure 6, our comparison (Kochkina et al., 2023) among various models, including branchLSTM (Kochkina and Liakata, 2020), TD-RvNN (Ma et al., 2018), BiGCN (Bian et al., 2020), SAVED (Dougrez-Lewis et al., 2021) and BERT (Devlin et al., 2019) for rumour detection evaluated on Twitter15, Twitter16 and PHEME (Kochkina et al., 2018), reveals there is no model that always performs the best. Although state-of-the-art models can achieve high accuracy on their training datasets, such performance drops quickly while evaluating on a different dataset (Kochkina et al., 2023). Due to the limitation of existing models in generalisation, users should interpret this result with caution as the system cannot guarantee output correctness.

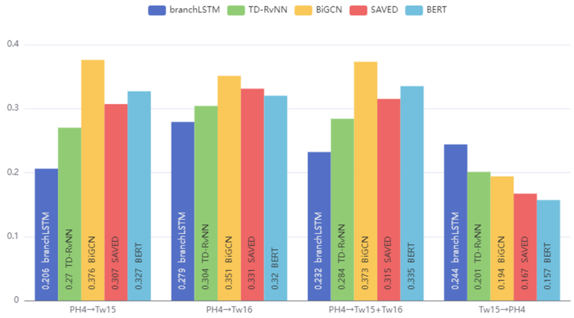


Figure 6: Cross-dataset evaluation of models train and test on different datasets, such as training on PHEME, testing on Twitter15/Twitter16 and vice versa.

[9] https://huggingface.co/digitalepidemiologylab/covid-twitter-bert-v2

5 Conclusion

This paper introduces a web-based system on *fact-checking* and *rumour detection* based on novel natural language processing models for COVID-19 misinformation detection. Going forward, we will keep updating the data and explore other methods for misinformation identification to improve the current system and introduce more functions to the system as part of our continuing efforts to support the general public to identify misinformation.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council (grant no. EP/V048597/1). YH and ML are each supported by a Turing AI Fellowship funded by the UK Research and Innovation (grant no. EP/V020579/1, EP/V030302/1).

References

Miguel Arana-Catania, Elena Kochkina, Arkaitz Zubiaga, Maria Liakata, Rob Procter, and Yulan He. 2022. Natural language inference with self-attention for veracity assessment of pandemic claims. In *Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1496–1511.

Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health Surveill*, 6(2).

Limeng Cui and Dongwon Lee. 2020. Coaid: Covid-19 healthcare misinformation dataset. *CoRR abs/2006.00885*.

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*, page 4171–4186.

John Dougrez-Lewis, Maria Liakata, Elena Kochkina, and Yulan He. 2021. Learning disentangled latent topics for twitter rumour veracity classification. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, page 3902–3908.

Samantha Finn, Panagiotis Takis Metaxas, and Eni Mustafaraj. 2014. Investigating rumor propagation with twittertrails. *CoRR abs/1411.3550*.

Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. 2020. Detecting covid-19 misinformation on social media. In *the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.

Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, SM Murshid Hasan, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, et al. 2020. Covid-19–related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, 103(4):1621.

Elena Kochkina, Tamanna Hossainb, Robert L.Logan IV Miguel Arana-Catania, Rob Procter, Arkaitz Zubiaga, Sameer Singh, Yulan He, and Maria Liakata. 2023. Evaluating the generalisability of neural rumour verification models. *Information Processing & Management*, 60(1).

Elena Kochkina and Maria Liakata. 2020. Estimating predictive uncertainty for rumour verification models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 6964–6981.

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour stance classification,detection and verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, page 3402–3413.

Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. 2020. Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. *CoRR abs/2011.04088*.

Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2020. Fine-grained fact verification with kernel graph attention network. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 7342–7351.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, page 1980–1989.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2020. Adversarial nli: A new benchmark for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 4885–4901.

Rodrigo Nogueira, Zhiying Jiang, Ronak Pradeep, and Jimmy Lin. 2020. Document ranking with a pretrained sequence-to-sequence model. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, page 708–718.

Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio Arechar, Dean Eckles, and David Rand. 2020. Understanding and reducing the spread of misinformation online. *NA - Advances in Consumer Research*, 48:863–867.

Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. 2020. Covid-19 on social media: Analyzing misinformation in twitter conversations. *CoRR abs/2003.12309*.

Lin Tian, Xiuzhen Zhang, and Jey Han Lau. 2022. Duck: Rumour detection on social media by modelling user and comment propagation networks. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*, pages 4939 – 4949.

Yariv Tsfati, H. G. Boomgaarden, J. Strömbäck, R. Vliegenthart, A. Damstra, and E. Lindgren. 2020. Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44(2):157–173.

Ellen Voorhees, Tasmeer Alam, Steven Bedrick, Dina Demner-Fushman, William R Hersh, Kyle Lo, Kirk Roberts, Ian Soboroff, and Lucy Lu Wang. 2021. Trec-covid: constructing a pandemic information retrieval test collection. In *ACM SIGIR Forum*.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020a. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *CoRR abs/2002.10957*.

Xuan Wang, Yingjun Guan, Weili Liu, Aabhas Chauhan, Enyi Jiang, Qi Li, , David Liem, Dibakar Sigdel, J. Harry Caufield, Peipei Ping, and Jiawei Han. 2020b. Evidenceminer: Textual evidence discovery for life sciences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, page 56–62.

Zilong Zhao, Jichang Zhao, Yukie Sano, Orr Levy, Hideki Takayasu, Misako Takayasu, Daqing Li, Junjie Wu, and Shlomo Havlin. 2020. Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, 9(7).

Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019. Gear: Graph-based evidence aggregating and reasoning for fact verification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, page 892–901.

Zhengyuan Zhu, Kevin Meng, Josue Caraballo, Israa Jaradat, Xiao Shi, Zeyu Zhang, Farahnaz Akrami, Fatma Arslan Haojin Liao, Damian Jimenez, Mohammed Samiul Saeef, Paras Pathak, and Chengkai Li. 2021. A dashboard for mitigating the covid-19 misinfodemic. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, page 99–105.