

# Document Retrieval and Claim Verification to Mitigate COVID-19 Misinformation

Megha Sundriyal<sup>1</sup>, Ganeshan Malhotra<sup>2</sup>, Md Shad Akhtar<sup>1</sup>, Shubhashis Sengupta<sup>3</sup>,  
Andrew Fano<sup>3</sup>, Tanmoy Chakraborty<sup>1</sup>

<sup>1</sup>IIT-Delhi, India, <sup>2</sup>BITS Pilani, Goa, India, <sup>3</sup>Accenture Labs, India

meghas@iiitd.ac.in, f20170512g@alumni.bits-pilani.ac.in, shad.akhtar@iiitd.ac.in,  
shubhashis.sengupta@accenture.com, andrew.e.fano@accenture.com, tanmoy@iiitd.ac.in

## Abstract

During the COVID-19 pandemic, the spread of misinformation on online social media has grown exponentially. Unverified bogus claims on these platforms regularly mislead people, leading them to believe in half-baked truths. The current vogue is to employ manual fact-checkers to verify claims to combat this avalanche of misinformation. However, establishing such claims' veracity is becoming increasingly challenging, partly due to the plethora of information available, which is difficult to process manually. Thus, it becomes imperative to verify claims automatically without human interventions. To cope up with this issue, we propose an automated claim verification solution encompassing two steps – document retrieval and veracity prediction. For the retrieval module, we employ a hybrid search-based system with BM25 as a base retriever and experiment with recent state-of-the-art transformer-based models for re-ranking. Furthermore, we use a BART-based textual entailment architecture to authenticate the retrieved documents in the later step. We report experimental findings, demonstrating that our retrieval module outperforms the best baseline system by 10.32 NDCG@100 points. We escort a demonstration to assess the efficacy and impact of our suggested solution. As a byproduct of this study, we present an open-source, easily deployable, and user-friendly Python API that the community can adopt.

## 1 Introduction

The escalating drift of online social media platforms has led to a massive rise in online content consumers. Participation in these platforms has swung into another correspondence, which is no longer limited by physical barriers. Because of their speed and focused information, these platforms facilitate the dissemination of personal thoughts and information to a much larger audience. However, at the same time, these platforms

have enriched an equally docile environment for malicious users to promulgate fake news, bogus claims, rumors and misinformation. There have been numerous cases where the propagation of malicious unverified content has influenced the entire society. One such concrete example is the 2016 Presidential Elections in the United States, which witnessed the alarming impact of false news, with many citizens swayed by a fraudulent website (Grave et al., 2018). Allcott and Gentzkow (2017) revealed that nearly 25% of American citizens visited a fake news website that aimed at manipulating the general public's cognitive process and consequently clouded the eventual conclusion of the election. Another recent example is the global pandemic of COVID-19. When the entire world went into lockdown, the virtual world encountered a great closeness transforming social media platforms into the primary conduits for information consumption and dissemination. Consequently, there has been an accretion of 50%-70% in total Internet hits in the year 2020 (Beech, 2020). Around the same time, enormous social media posts with unverified bogus claims about the pandemic began to arise, frequently spurring life-threatening remedies (Naeem and Bhatti, 2020). Such claims had an unprecedented impact, resulting in monetary damage and the loss of priceless human lives. A study revealed that at least 800 individuals died worldwide in the first quarter of 2020 due to misinformation about COVID-19 (Coleman, 2020).

**Motivation:** A slew of such incidents has continued to emerge from the worldwide community in recent years. Thousands of people read these unverified claims online and spread misinformation if the claims' integrity is not corroborated. As a result, a variety of manual fact-checking organizations have evolved to

address this concerning issue. Unfortunately, the enormity of misinformation floating around on the Internet has developed into a global *infodemic*<sup>1</sup> making their efforts untenable. To alleviate this bottleneck, the process of automating fact-checking has recently garnered a lot of consideration in the research world. Vlachos and Riedel (2014) formalized the task of fact-checking and claim verification as a series of components – identifying claims to be evaluated, extracting relevant shreds of evidence, and delivering verdicts. As a result, this facilitated the establishment of automated fact-checking pipelines composed of subcomponents that can be mapped to tasks well-studied in the NLP community. The task of retrieving relevant information has gained a lot of impetus in recent years, especially with the introduction of tools like PYSERINI<sup>2</sup> and BEIR<sup>3</sup>. Furthermore, advancements were made by establishing datasets of either claims acquired from fact-checking websites (Wang, 2017) or datasets curated specifically for research (Thorne et al., 2018a). The recent release of the CORD-19 dataset<sup>4</sup>, consisting of more than 500,000 articles, has provided access to thousands of scientific articles on the prevention techniques, spread, transmission, and cures of the COVID-19. The dataset consists of more than 500,000 articles.

**State-of-the-art and Challenges:** Previous research in the realm of claim verification and fact-checking has primarily concentrated on structured data, often in the form of subject-predicate-object statements (Dong et al., 2015; Nakashole and Mitchell, 2014). Several research on detecting false claims on social media included network metadata such as user profile characteristics, user-user interactions, popularity attributes based on the number of likes or followers, etc (Kumar et al., 2016; Qazvinian et al., 2011). Most notably, all of these procedures use black-box approaches, and hence, do not articulate why a statement is considered verified. Another pressing issue is that the input claim does not coexist naturally with the corresponding review articles. As a result, obtaining the relevant articles via internet

is critical. There is, however, a disparity between the human—crafted review articles generated specifically for claim verification in the fact database and the report articles gathered from the web. Meanwhile, methods such as ClaimBuster<sup>5</sup> and Google’s Fact Check Explorer<sup>6</sup> have been developed to check the legitimacy of the statement by assessing trust criteria utilizing internet. However, these existing methods are not intended to investigate the veracity of the evidence and hence fail to meet the previously identified issues.

**Our Contributions:** To address the aforementioned issues, we create an end-to-end claim verification system capable of establishing the integrity of a query claim and explaining its decisions with supporting evidence. Our model takes in as input the claim whose veracity is to be verified. Due to the diversity of natural language idioms, the first major problem in developing such a system is identifying connected snippets of a claim. Thus, we utilize well-known retrieval systems for this task. The system selects relevant articles from either the CORD-19 dataset or our in-house dataset, *ClaveR*, using a host of different models ranging from BM25 to intricate hybrid searchers. Users can additionally opt to retrieve more fine-grained results where the model selects relevant snippets in the article. Eventually, the model verifies the claim by calculating the entailment of the input claim concerning the retrieved articles.

Through this work, we make the following contributions:

1. To allay the unavailability of a COVID-19 centric annotated dataset for claim verification in Twitter, we develop *ClaveR*, a new dataset of claim-evidence pairs based on a subset of COVID-19-related claims reaped from a recently released large-scale claim-detection dataset, LESA (Gupta et al., 2021).
2. We propose an end-to-end claim verification system encompassing two steps to validate the claims proffered online provided high-quality editorial review articles and Twitter posts.

<sup>1</sup><https://www.who.int/health-topics/infodemic>

<sup>2</sup><https://github.com/castorini/pyserini>

<sup>3</sup><https://github.com/UKPLab/beir>

<sup>4</sup><https://allenai.org/data/cord-19>

<sup>5</sup><https://idir.uta.edu/claimbuster/api/>

<sup>6</sup><https://toolbox.google.com/factcheck/explorer>

3. We evaluate our retrieval model against multiple state-of-the-art systems concerning our dataset, *ClaveR*. According to the comparison, BM25 surpasses all other existing systems by a wide margin.
4. We provide an open-source, easily deployable, and user-friendly Python API based on our proposed solution for claim verification. We also accompany a demonstration to evaluate the efficacy and usage of the API.

## 2 Related Work

The challenge of verifying claims on online social media has garnered considerable attention in the last several years. Initially, the task of automatic claim verification and fact-checking were investigated in the context of computational journalism (Cohen et al., 2011; Flew et al., 2012), and journalists and professional fact-debunkers manually verified claims utilizing various information sources. However, that was not just time-consuming but also introduced substantial human bias in it. The recent advancement in NLP and information retrieval (IR) has equipped journalists and online social media users with tools enabling automatic claim verification. In the past few years, plenty of work has been proposed to fact-check online claims. Vlachos and Riedel (2014) presented the initial pioneering work in this domain. They published the first claim verification dataset, which included 106 statements taken from fact-checking websites like PolitiFact. However, they lacked justification for the verdict, which verification systems typically require. To address this issue, Wang (2017) prolonged this approach by introducing 12.8K claims from PolitiFact along with their explanations. The Fact Extraction and Verification (FEVER) shared task was launched to advance research in this direction (Thorne et al., 2018b). The organizers of the FEVER shared task constructed a large-scale dataset of 185445 claims based on Wikipedia articles, each of which comes with several evidence sets.

Traditionally, the existing claim verification systems primarily rely on textual content and/or social context. The content-based methods essentially acquire the n-grams (Wang, 2017), semantics (Khatter et al., 2019), writing styles (Gröndahl and Asokan, 2019), etc. Besides textual-content, auxiliary knowledge around social-context has also been extensively examined for verification tasks.

These context-based methods emphasize collecting user profile-based (Shu et al., 2019), propagation structure-based (Wei et al., 2019), source-based (Pennycook and Rand, 2019), etc. Zhi et al. (2017) introduced ClaimVerif that provides a credibility score for a user given a claim and also gives supporting evidences that justify the credibility score. Hanselowski et al. (2018) presented their approach to the FEVER task (Thorne et al., 2018b) which was introduced to expedite the development of fact verification systems, in which they used entity linking for document retrieval and Enhanced Sequential Inference Model for determining the entailment. Ma et al. (2019) used Hierarchical Attention Networks with sentence-level evidence embeddings. Despite the fact that these tactics produce good performance results, it is challenging for these approaches to provide adequate reasons for claim verification outcomes.

As a result, current research has focused on interpretable claim verification, which develops interactive models to examine the distinction. Attention-based interaction models (Popat et al., 2018), gate fusion interactive models (Wu and Rao, 2020), coherence modelling interactive models (Ma et al., 2019), and graph-aware interaction models are among the interactive models. The granularity of captured semantic conflicts involves word-level (Popat et al., 2018), sentence-level (Ma et al., 2019), and multi-feature (Wu and Rao, 2020) conflicts. Su et al. (2020) came up with a question-answering-based model that mines relevant articles from the CORD-19 dataset and summarizes them to answer pressing questions about the COVID-19 pandemic. Recently, Pradeep et al. (2021) proposed a T5<sup>7</sup> transformer-based architecture for abstract retrieval, sentence selection and label prediction and perform claim verification. Similar to us, they also utilized the CORD-19 (Wang et al., 2020) corpus as the knowledge base to retrieve shreds of evidences. These methods, which employ semantic conflicts to verify claims, reflect a certain degree of interpretability. But not all conflicts can be used as valid evidence to reasonably explain the results, and they also include considerable conflicts unrelated to claims or even interfere with the verified results. It is difficult for automatic claim verification to provide reasonable explanations for the

<sup>7</sup>[https://huggingface.co/transformers/model\\_doc/t5.html](https://huggingface.co/transformers/model_doc/t5.html)

Table 1: Examples from ClaVer dataset along with the evidence and corresponding labels.<sup>8</sup>

<b>Claim: 1</b>	
<i>@CNN Boosting our immune systems will help deter the virus. It's our only defense aside from n95 masks and goggles</i>	
<b>Evidence</b>	<b>Label</b>
<i>First, there's the not-so-great news. Despite claims you may have seen on the Internet, there's no magic food or pill that is guaranteed to boost your immune system and protect you against coronavirus...<b>There are ways to keep your immune system functioning optimally, which can help to keep you healthy and give you a sense of control in an uncertain time...</b>For a starter dose of immune-boosting vitamins, minerals and antioxidants, fill half of your plate with vegetables and fruits.</i>	<b>SUPPORTED</b>
<b>Claim: 2</b>	
<i>@AFP @EvelDick It's much more than a coincidence that China has a bioweapons lab with sloppy protocols in Wuhan. Wonder if this is another booboo? Seems like a very bad place to have a bioweapons lab. The whole "this came from snakes" Chinese party line makes me think the virus was manufactured.</i>	
<b>Evidence</b>	<b>Label</b>
<i>As the Covid-19 pandemic continues its destructive course, two theories are being widely aired...<b>The lab is one of 20 such facilities under the Chinese Academy of Sciences, but is the only one dealing with virology. Fully compliant with ISO standards, the Wuhan facility interacts regularly with a host of outside experts. Like other labs, its aim is to protect populations against new viruses...</b></i>	<b>REFUTED</b>

verification results; the demand for interpretable claim verification is growing, with the goal of providing end-users with grounds to debunk rumours by showing the incorrect elements of claims. Existing methods in this assignment investigate semantic conflicts between claims and relevant articles by creating various interactive models to explain verification results.

### 3 Description of the Datasets

For our experiments, we adopt two datasets. Their details are shown as follows:

1. **CORD-19 Dataset** (Wang et al., 2020): CORD-19 dataset consists of over  $\sim 500,000$  articles (over  $\sim 200,000$  containing full text) taken from various scientific publications about COVID-19, SARS-COV2 and other viruses. This dataset provides access to trustworthy scientific sources of information to mitigate the spread of misinformation.
2. **LESA Dataset** (Gupta et al., 2021): LESA dataset consists of  $\sim 10,000$  tweets that were mined from various sources and were manually annotated for the binary classification task of claim detection. Furthermore, we develop a validation set – **Claim Verification (ClaVer)** by selecting a subset of claims from the LESA dataset and annotating those claims with relevant articles that provide additional context for the claim, as shown in Table 1. These articles are gathered from reliable online news sources

and contain additional extensive information that may be used to verify the authenticity of the claim. The articles can “*Refute*” or “*Support*” the claim. In other circumstances, the claim may be that the annotated article does not give conclusive evidence. These articles lack sufficient information to support or reject the claim’s veracity and hence labelled for “*Not Enough Information*”. These articles are also stored in our global knowledge base of articles along with the articles taken from the CORD-19 dataset.

### 4 Our Approach

Adhering to the standard of automated claim verification and fact-checking systems (Thorne et al., 2018b), our proposed approach also consists of a two-step pipeline – Document Retrieval and Veracity Prediction. In this section, we present the techniques employed for retrieval and veracity prediction components. Besides the current approach, we had also employ alternative techniques using Rapid Automatic Keyword Extraction or RAKE (Rose et al., 2010) and SciSpacy (Neumann et al., 2019) for keyword extraction and searching our corpus using the extracted keywords. Figure 1 illustrates the general architecture of our proposed claim verification approach. Once a textual claim is submitted, the document retrieval module extracts the top-k relevant documents from the knowledge base. The retrieved documents are then passed to the veracity prediction module that figures out an



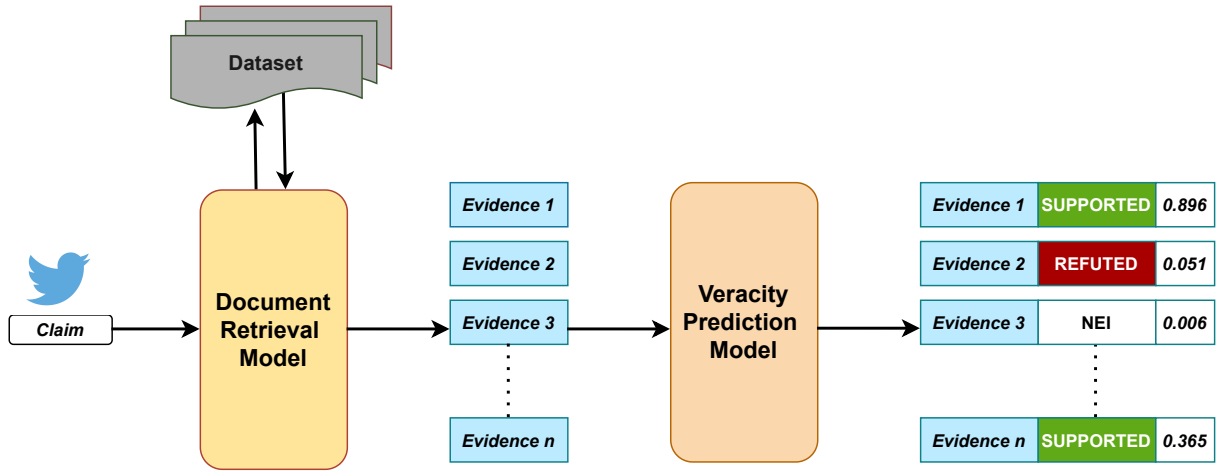


Figure 1: An overview of the proposed evidence-based claim verification pipeline. The significant components have been highlighted to correspond to the two stages of our experimental setup: (a) a document retrieval module that uses one of the given datasets to retrieve top-k relevant documents for the corresponding input claim, and (b) a veracity prediction module that seeks to establish the retrieved documents’ credibility against the input claim.

entailment decision for the claim with respect to the retrieved evidences.

#### 4.1 Document Retrieval

Inspired by IR systems, the retrieval problem we attempt to address is defined as follows: Given a textual claim  $c$  and a set of documents  $D$ , we aim to retrieve the top-k documents from  $D$  relevant to  $c$ . Our retrieval pipeline consists of two broad categories of retrieval systems, namely Sparse Retrieval and Dense Retrieval.

1. **Sparse Retrieval Model:** Over the years, lexical approaches like TF-IDF and BM25 have dominated textual information retrieval. We also utilize the BM25 scoring function (Robertson et al., 1995) as the backbone model for sparse retrieval. We use the sparse retrievers for both the ClaVer as well as CORD-19 datasets. In this case, we also provide an extra option of getting finer-grained results. This step scans through the retrieved article and provides a relevant part of the article. We use a BioBERT (Lee et al., 2019) language model which is pre-trained on large-scale bio-medical corpora. We compute the hidden representation of each paragraph in the article using the language model and calculate its cosine similarity with the hidden representation of the claim. The paragraph with the highest value is then selected.
2. **Dense and Hybrid Retrieval Models:** More recently, dense retrieval approaches were pro-

posed to get better retrieval results. They are capable of capturing semantic matches and try to overcome the (potential) lexical gap. Dense retrievers map queries and documents in a shared, dense vector space (Gillick et al., 2018). This allowed the document representation to be pre-computed and indexed. We provide the option of dense retrievers specifically for our ClaVer dataset. Using dense indexes for CORD-19 dataset is difficult because of the huge size of the corpora. To use the dense and hybrid searchers, we first index our ClaVer data using the FAISS (Johnson et al., 2017) library. For our dense retriever, we use the simple dense searcher provided by the PYSERINI (Lin et al., 2021) library while initializing it with COVID-BERT weights. The hybrid searcher uses a combination of sparse and dense retrievers and computes a weighted interpolation of the individual results to arrive at the final rankings. We use the TCT-ColBERT (Lin et al., 2020) architecture to encode our queries into the same representation space as the encoded documents.

#### 4.2 Veracity Prediction

Given a claim and the evidence gathered through document retrieval system, veracity prediction module seeks to establish the evidence’s credibility in terms of a veracity score. To verify the veracity of our retrieved articles, we leverage a BART-based

Table 2: Sample response generated by our proposed system leveraging ClaVer dataset for extraction.

Claim			
Story about how #HydroxyChloroquine likely help people recover from #Coronavirus. IMO, it was never touted as the cure but as option for treatment doctors should consider and it appears to work in some cases....39 in one place.			
Outputs			
Technique	Evidence Retrieved <sup>8</sup>	Label	Veracity
Ours	<i>Chloroquine and hydroxychloroquine, a pair of old drugs used to treat and prevent malaria, are the latest compounds to be thrust into the limelight as people tout them as treatments for the novel coronavirus. On Sunday, March 29, the US Department of Health and Human Services accepted 30 million doses of hydroxychloroquine sulfate from Novartis and 1 million doses of chloroquine phosphate from Bayer...The World Health Organization is sponsoring a large international clinical trial called SOLIDARITY to study six drugs that could be rapidly deployed for the fight the coronavirus, including chloroquine and hydroxychloroquine.</i>	CONTRADICTION	0.82737
Dense	<i>As of now, no study says coronavirus can be cured by drinking lots of water or gargling with warm saltwater. Though it is true that warm salt water has long been used as a home remedy to soothe a sore throat, but till now, there is no evidence that it can also ward off the novel coronavirus. A report by fact-check website "Snopes" also says that there is no proof that coronavirus remains in the throat for four days as mentioned in the viral post.</i>	NEUTRAL	0.99825
Hybrid	<i>As of now, no study says coronavirus can be cured by drinking lots of water or gargling with warm saltwater. Though it is true that warm salt water has long been used as a home remedy to soothe a sore throat, but till now, there is no evidence that it can also ward off the novel coronavirus. A report by fact-check website "Snopes" also says that there is no proof that coronavirus remains in the throat for four days as mentioned in the viral post.</i>	NEUTRAL	0.99825

(Lewis et al., 2020) Natural Language Inference (NLI) model that returns one of the three classes for each claim-evidence pair: Entailment, Neutral and Contradiction (as shown in Table 2). The mapping of these labels with our use case is done in the following way:

- If the model outputs ‘Entailment’, it means that the given claim’s veracity can be positively supported by the retrieved article.
- If the model outputs ‘Contradiction’, it means that the given claim’s veracity is refuted by the retrieved article which makes the claim dubious.
- If the model outputs ‘Neutral’, it means the retrieved article does not provide enough evidence to either support or refute the claim.

## 5 Evaluation

We compare the findings of our retrieval system BM25 to those of other existing systems. We employ a collection of claims and ground-truth labels from our ClaVer dataset for quantitative evaluation. The test data set consists of claims excluded from the knowledge base in the retrieval phase. For this, we develop a manually annotated dataset with  $\sim 1000$  claims obtained from Twitter and build a knowledge-base of  $\sim 400$  articles from reliable sources, equipping a testing ground to validate the results. Table 3 presents experimental results based on Normalized Discounted Cumulative Gain (NDCG@k) scores, Mean Average Precision (MAP@k) and Mean Average Recall (MAR@k) scores for different values of k. We find that using BM25 outperforms all other baseline systems for retrieval task. The NDCG@100 score of the BM25

Table 3: Performance of various retrieval techniques on ClaVer dataset. (NDCG: Normalized Discounted Cumulative Gain, MAP: Mean Average Precision and MAR: Mean Average Recall)

Technique	NDCG@1	NDCG@10	NDCG@100	MAP@1	MAP@10	MAR@1	MAR@10
Ours	<b>24.71</b>	<b>36.75</b>	<b>45.73</b>	<b>24.71</b>	<b>32.14</b>	<b>24.71</b>	<b>51.72</b>
CrossEncoder MS Marco	22.99	35.41	35.41	22.99	31.12	22.99	48.85
CrossEncoder CovidBERT	3.41	15.04	15.04	3.41	3.41	3.49	36.36
SentenceBERT MS Marco	18.97	32.09	32.58	18.97	26.83	18.97	49.43

Vaccines are not effective against COVID-19

**Evidence 1**

Link of the Document: <https://doi.org/10.1080/21645515.2020.1735227>

COVID-19, an emerging coronavirus infection advances and prospects in designing and developing vaccines, immuno...

CORD-19
BM25

The article belongs to neutral category with a confidence of 0.9328

**Evidence 2**

Link of the Document: <https://www.who.int/news-room/9-3-detail/herd-immunity/lockdowns-and-covid-19>

Herd immunity, also known as 'population immunity, is the indirect protection from an infectious d...

CLAVER
BM25

The article belongs to neutral category with a confidence of 0.9903

**Evidence 3**

Link of the Document: <https://oypost.com/2020/01/29/no-corona-virus-isnt-linked-to-corona-beer-cantbe-cured-with-bleach/>

Despite the catchy name, no,....

CLAVER
Dense Search

The article belongs to neutral category with a confidence of 0.9982

**Evidence 4**

Link of the Document: <https://oypost.com/2020/01/29/no-corona-virus-isnt-linked-to-corona-beer-cantbe-cured-with-bleach/>

Despite the catchy name, no, coronavirus has nothing to do...

CLAVER
Hybrid Search

The article belongs to neutral category with a confidence of 0.9982

Figure 2: User-interface of our proposed tool after the claim has been submitted.

retrieval model improves the baseline method by more than 10% out of the whole testing set. We find that BM25 detects relevant snippets with higher precision and recall than other existing retrieval systems.

## 6 Demonstration

In this section, we demonstrate how our proposed claim verification pipeline works. Figure 2 depicts an example claim as well as the model’s output results. Users enter a claim into our system as a query, and the system evaluates whether or not it is a validated claim. In practice, the system takes somewhere around 20 and 80 seconds to execute a single user query, depending on the number and length of articles obtained by the search engine.

The input section of our tool, as shown in Figure 2, provides a query text box where the user can enter any natural language text as an input claim for evaluation, as well as a specific configuration to

limit the number of articles to be retrieved. Following the submission of the claim, the tool’s back-end server does its analysis. It returns three sets of outputs: (i) a set of articles employing the various approaches, (ii) a claim category, and (iii) a veracity score. The output also presents the technique utilized for retrieval (pink) and from which knowledge base the shreds of evidence were extracted (blue). The most intriguing aspect of the system is that it links resources from the web, where the article was retrieved, allowing individuals to make their own decisions based on them.

Not all information is equally reliable, and sometimes even the trusted sources contradict one another. This calls into question the assumptions behind most current fact-checking research, which relies on a single authoritative source. As a result, we offer results for a common claim from several models and knowledge bases. For demonstration, we practice the widely spread claim “*Vaccines are not effective against COVID-19*” as an input as shown in Figure 2, and the tool returned the top-

<sup>8</sup>Links of article sources can be found at: <https://cutt.ly/1FwsxXa>

ranked shreds of evidence. The first two pieces of evidence come from the BM25 model, which was run on the CORD-19 dataset and our data, respectively. Furthermore, evidences 3 and 4 collected articles from our dataset using a dense and hybrid retrieval strategy, respectively. We can see that all four pieces of evidence assigned the same label to the claim, but their truthfulness scores differed from each other.

## 7 Conclusion

In this work, we verged upon claim verification on online social media towards coping with misinformation. We bestowed a claim verification system that evaluates the authenticity of a user-supplied query claim and justifies the verdict corroborating evidence. We explored multiple retrieval methodologies and published user research findings, demonstrating the utility of the BM25 method. Unlike other tools, our system learns the distributed representations to encapsulate the semantic relations between the claim and the evidence. Our approach uses a two-step training process to provide a high-quality veracity score as well as best-suited articles, leveraging data from formal articles and web-based informal texts. We have made the source codes and the dataset public at the following link: [https://github.com/LCS2-IIITD/claim\\_verification](https://github.com/LCS2-IIITD/claim_verification).

## Acknowledgements

T. Chakraborty would like to acknowledge the support of the Ramanujan Fellowship, and ihub-Anubhuti-iiitd Foundation set up under the NM-ICPS scheme of the Department of Science and Technology, India. M. S. Akhtar and T. Chakraborty thank Infosys Centre for AI at IIIT-Delhi for the valuable support.

## References

- Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.
- Mark Beech. 2020. [Covid-19 pushes up internet use 70% and streaming more than 12%, first figures reveal](#).
- Sarah Cohen, Chengkai Li, Jun Yang, and Cong Yu. 2011. Computational journalism: A call to arms to database researchers. In *5th Biennial Conference on Innovative Data Systems Research, ACM*.
- Alistair Coleman. 2020. [‘hundreds dead’ because of covid-19 misinformation](#).
- Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: Estimating the trustworthiness of web sources.
- Terry Flew, Christina Spurgeon, Anna Daniel, and Adam Swift. 2012. The promise of computational journalism. *Journalism practice*, 6(2):157–171.
- Daniel Gillick, Alessandro Presta, and Gaurav Singh Tomar. 2018. End-to-end retrieval in continuous space. *arXiv preprint arXiv:1811.08008*.
- Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*.
- Tommi Gröndahl and N Asokan. 2019. Text analysis in adversarial settings: Does deception leave a stylistic trace? *ACM Computing Surveys (CSUR)*, 52(3):1–36.
- Shreya Gupta, Parantak Singh, Megha Sundriyal, Md Shad Akhtar, and Tanmoy Chakraborty. 2021. Lesa: Linguistic encapsulation and semantic amalgamation based generalised claim detection from online content. In *Proceedings of the 16th Conference of the EACL: Main Volume*, pages 3178–3188.
- Andreas Hanselowski, Hao Zhang, Zile Li, Daniil Sorokin, Benjamin Schiller, Claudia Schulz, and Iryna Gurevych. 2018. UKP-athene: Multi-sentence textual entailment for claim verification. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, Brussels, Belgium. ACL.
- Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2017. Billion-scale similarity search with gpus. *arXiv preprint arXiv:1702.08734*.
- Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, pages 2915–2921.
- Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on World Wide Web*, pages 591–602.
- Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training



- for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the ACL*, Online. ACL.
- Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira. 2021. [Pyserini: An easy-to-use python toolkit to support replicable ir research with sparse and dense representations](#).
- Sheng-Chieh Lin, Jheng-Hong Yang, and Jimmy Lin. 2020. [Distilling dense representations for ranking using tightly-coupled teachers](#).
- Jing Ma, Wei Gao, Shafiq Joty, and Kam-Fai Wong. 2019. Sentence-level evidence embedding for claim verification with hierarchical attention networks. ACL.
- Salman Bin Naeem and Rubina Bhatti. 2020. The covid-19 ‘infodemic’: a new front for information professionals. *Health Information & Libraries Journal*, 37(3):233–239.
- Ndapandula Nakashole and Tom Mitchell. 2014. Language-aware truth assessment of fact candidates. In *Proceedings of the 52nd Annual Meeting of the ACL (Volume 1: Long Papers)*, pages 1009–1019.
- Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. 2019. [ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing](#). In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 319–327, Florence, Italy. ACL.
- Gordon Pennycook and David G Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7):2521–2526.
- Kashyap Papat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. Declare: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint arXiv:1809.06416*.
- Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira, and Jimmy Lin. 2021. Scientific claim verification with VerT5erini. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, online. ACL.
- Vahed Qazvinian, Emily Rosengren, Dragomir 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.
- Stephen E Robertson, Steve Walker, Susan Jones, Micheline M Hancock-Beaulieu, Mike Gatford, et al. 1995. Okapi at trec-3. *Nist Special Publication Sp*, 109:109.
- Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. *Automatic Keyword Extraction from Individual Documents*, chapter 1. John Wiley Sons, Ltd.
- Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 312–320.
- Dan Su, Yan Xu, Tiezheng Yu, Farhad Bin Siddique, Elham J. Barezi, and Pascale Fung. 2020. [Caire-covid: A question answering and query-focused multi-document summarization system for covid-19 scholarly information management](#).
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018a. Fever: a large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355*.
- James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. 2018b. The fact extraction and VERification (FEVER) shared task. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, Brussels, Belgium. ACL.
- Andreas Vlachos and Sebastian Riedel. 2014. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, Baltimore, MD, USA. ACL.
- Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darin Eide, Kathryn Funk, Yannis Katsis, Rodney Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex Wade, Kuansan Wang, Nancy Xin Ru Wang, Chris Wilhelm, Boya Xie, Douglas Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. 2020. [Cord-19: The covid-19 open research dataset](#).
- William Yang Wang. 2017. “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the ACL (Volume 2: Short Papers)*, Vancouver, Canada. ACL.
- Penghui Wei, Nan Xu, and Wenji Mao. 2019. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. *arXiv preprint arXiv:1909.08211*.
- Lianwei Wu and Yuan Rao. 2020. Adaptive interaction fusion networks for fake news detection. *arXiv preprint arXiv:2004.10009*.
- Shi Zhi, Yicheng Sun, Jiayi Liu, Chao Zhang, and Jiawei Han. 2017. Claimverif: a real-time claim verification system using the web and fact databases. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2555–2558.