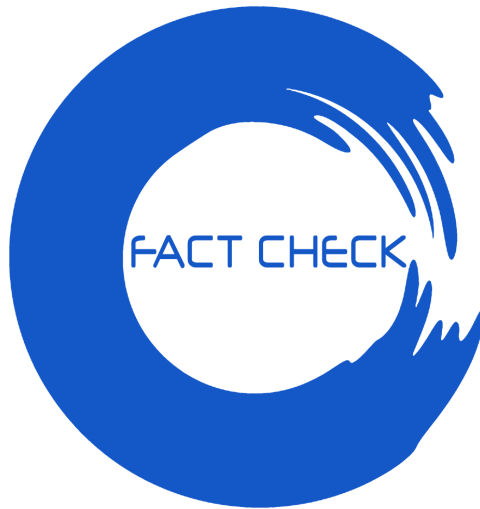# Fact Check

AI Fact Checking and Claim Correcting System

Han Lee, Sukhamrit Singh
Stanford Compression Forum

STEM To SHTEM 2021

Abstract

The creation of the internet sparked an age where information is available with the click of a button. Humans all around the world have been able to use this resource to their benefit. Recipes, travelling guides, political stances, anything that can be thought of can be found in this expansive cloud. With this rise however, misinformation runs rampant as false information is spread with the potential to persuade users. As a response to this, our group has developed an AI Fact Checking and Claim Correcting System that checks claims for their validity. Our project makes use of the Transformers NLP which utilizes a large database to cross reference the information that is inputted. In order to score the validity of a claim, our project uses the DistilBERT and RoBERTa AI models. The models try to predict words which fit best in the context of the claim, and the more similar the predicted claim is from the original claim, the higher accuracy rating the original claim gets. After scoring a claim, the Sentence Correction System corrects the original claim by replacing each word in it with the top predictions from the models. The corrected claim is then passed into the Wikipedia Tokenizer System which converts the words from the corrected claim into tokens, and searches each token on Wikipedia. The data from the Wikipedia articles are stored and used to create a knowledge graph for the corrected claim. Our AI Fact Checking and Claim Correcting System is also extended to a mobile app called "Fact Check," where the user can input their claim as a spoken query and the accuracy results will be shown.

Keywords: NLP, DistilBERT, RoBERTa, Sentence Correction System, Wikipedia Tokenizer System

1 Introduction

1.1 Natural Language Processing (NLP)

Natural language processing is used to emulate human language. Through the use of context and repeated patterns, NLP attempts to correctly predict words that fit best within the context of the sentence and can resemble our language to the best of its abilities. The research for this form of AI began in the 1940s when Weaver and Booth began to develop Machine Translation (MT) [7]. The format behind MT was very textbook and only referred to the dictionary in order to replace words. It did not take long for this approach to produce poor results. Later in 1957, Chomsky proposed the idea of generative grammar which stated that all sentences have a logical reasoning behind them

[7]. This opened a whole new perspective on the topic which continued on till now. In the past decade, work on NLP has developed rapidly resulting in many innovations.

1.2 Knowledge Graphs

Knowledge graphs are an immense collection of data that are used for making connections between interrelated-subjects. These connections are formed through the use of nodes, where each node is connected to other nodes carrying information on similar topics. To create these connections, large data sources, such as Wikipedia, are scavenged through to collect as much information on a specific subject. Thus, based on what is being searched for, node A may be connected to node B for the similarities that they share. The multiple connections that can be formed between nodes eventually creates a tree data structure, where a knowledge graph can contain multiple of these structures.

If the sentence "Barack Obama is the 44th president of the USA" is inputted to create a basic knowledge graph, the program would extract the key components of the sentence, such as "Barack Obama", "44th president", and "USA." With these tokenized components, a simple graph is created:
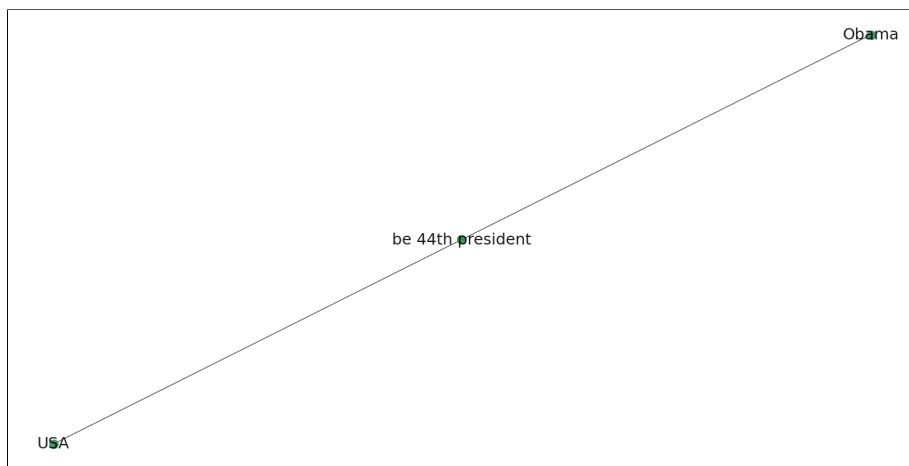

Fig. 1 An example of a knowledge graph

1.3 Fact Checking

Fact checking is the process in which statements are checked for their validity. Once the whole of a claim is checked, it is classified on whether it is factually accurate or not. Traditionally, humans have been the ones to execute this process. Because of the amount of work required to complete such an arduous task, the process can take anywhere from a day to weeks. In order to lighten

this load, the idea of using AI in this process has been introduced. However, this new source of technology is still in its development phases, meaning it does not have the complete knowledge to accurately fact check claims covering a vast array of topics. Thus, when the AI tends to make errors, humans still must be aware to correct them.

Fact checking is done through many different components such as quote verification, title verification, position claims, and the most important method, by triples [14]. The triple refers to subject, object and predicate. These components of a sentence are used to cross reference information from a data set to verify information. We used the method of checking with triples in our project to judge which parts of a sentence were greater in significance and which parts would provide the most accurate validity ratings.

2 Related Works

Due to the rapid rise in the spread of misinformation, NLPs have seen continued use in the field of fact checking. Our group specifically used the Transformers NLP throughout the entirety of our project. The Transformers NLP makes use of one of the most well known inputs for NLPs called triples. Another popular input that was targeted were textual claims. This input was the most important because of the role it plays in spreading misinformation. However, one difficulty that we as a group and other researchers have come upon is the trouble that the NLP has with proper nouns included inside these claims. Names are often hard to generate around because names can be shared by many people around the world. This fact makes it difficult for the NLP to correctly predict words.

Another aspect of our project that we shared with researchers was the use of knowledge graphs. When using NLPs, it is essential to provide a large amount of data upon which cross referencing can be done. Knowledge graphs, as seen in Vlacho and Riedel's research [16], help in the process of retrieving information when verifying the claim at hand with said information.

Lastly, the NLPs that we used during our project were works that were created before us. They fit best for our project since they were made to predict words that are taken out of a claim. We used the DistilBERT and RoBERTa-Large models to predict numerical claims, general claims, and proper nouns.

DistilBERT is a mask-filling-model from the Transformers NLP. This is one of many pre-trained models used to predict words based on the context of the sentence. DistilBERT was made to be a compact version of the original BERT model while retaining the same prediction values as the original model. Thus, this means that the DistilBERT model returns similar outputs at a faster rate due to its smaller size. The DistilBERT model was one of the models used in

our project.

RoBERTa-Large (a larger version of the RoBERTa model) is another pre-trained mask-filling-model from the Transformers NLP. This model is generally much smarter than the original BERT model. This is because it was pre-trained with an enormous collection of English data. As a whole, RoBERTa-Large's capabilities surpass those of BERT. The RoBERTa-Large model was another model used in our project.

3 Methods and Materials

3.1 Mask-Filling-Model

3.1.1 "Mask"

The word "mask" is a crucial keyword and a very important aspect of our project. What it does is it lets the program know where the AI model should predict a word that most accurately fits the context of the sentence or claim.

For example, if the mask-filling-model is to complete the claim "David Beckham played soccer, a sport also referred to as <mask>," the "mask" keyword in this case tells the model to predict the last word of the sentence. Based on the context of the claim, using its pre-trained knowledge, the AI model will predict 5 words which best fit the sentence:

Predicted Word	Confidence Score
football	0.5795354247093201
rugby	0.21709421277046204
basketball	0.05613641440868378
tennis	0.03103862702846527
soccer	0.022038333117961884

Fig. 2 Actual predictions from the model for the input: "David Beckham played soccer, a sport also referred to as <mask>."

The way the mask-filling-model outputs its predictions is from the order of most confident to least confident. We can see that the first prediction is accurate and is the answer we want, but the rest are either completely irrelevant or repeated words. However, the AI model deserves some credit for the rest of the predictions since they are sports related, indicating that the model used the context of the sentence to determine that David Beckham is an athlete.

If our program was to always take the prediction with the highest rating and add it to the original claim, our output would be "David Beckham played

soccer, a sport also referred to as football," which in the end, is a factually correct claim.

3.1.2 "DistilBERT vs. RoBERTa"

	DistilBERT	RoBERTa
Size	Base: 66	Base: 110 Large: 340
Training Time	4 times less than BERT	4-5 times more than BERT
Performance	3 % degradation from BERT	2-20 % improvement over BERT
Data	16 GB data	160 GB data

Fig. 3 Data from [12]

Why are both of the models being used in our AI system? When it came to testing which model was best at giving the most accurate validity ratings, each model was better at some aspect than the other. For example, DistilBERT was better at predicting numbers based on the context of the claim whereas the RoBERTa-Large model was better at predicting general information and facts. With this in mind, we thought that using the knowledge from both would be most beneficial to our project. Utilizing both the AI models will allow our program to determine a validity rating for a claim with greater accuracy.

3.2 Accuracy Rating system

In order to judge the accuracy of a claim being inputted into our program, a system had to be created where each word is graded on how well it fits within the context of the sentence. The way this was done was by replacing each word from the input with the "mask" keyword. The mask-filling-models were then called to predict which words would best fit within the context of the sentence. If any of the predicted words matched the original word, a score would then be assigned for that word.

The max score any word can achieve is the rating 1. However, this is only possible if the model predicts the same exact word and it is the highest rated prediction. If the prediction were to be the second strongest prediction instead, 0.2 will be subtracted by the score 1. Essentially, the rating is given by multiplying how many places the prediction is from the top prediction by 0.2 and subtracting that from 1. If the word from the original claim is never

predicted, it is given a rating of 0.

$$WordRating = 1 - (0.2 \times NumberOfPlacesFromTopPrediction) \quad (1)$$

Original Claim: "The 2020 Olympics were held in Japan."

Inputted Claim: "The 2020 [MASK] were held in Japan."

Predicted Word	Confidence Score
championships	0.26722633838653564
games	0.24994230270385742
paralympics	0.08590848743915558
olympics	0.05416667088866234
finals	0.04430991783738136

The word olympics is 3 places below the top predicted word

$$Rating = 1 - (0.2 \times 3) = 0.4 \quad (2)$$

Fig. 4 Scoring representation for one word (using the DistilBERT model)

After the AI models have gone through each word, all the scores are added together and divided by the total number of words in the claim (special characters also count as individual words).

$$ClaimAccuracy = \frac{TotalScoresOfAllWords}{TotalNumberOfWords} \quad (3)$$

Once the final rating is determined, it must pass an accuracy threshold, which is 0.85. If the total score for a given claim is 0.85 or above, the claim will be deemed "factually accurate," and if the rating is below this threshold, the claim will be deemed "factually inaccurate."

3.3 Sentence Correction Wikipedia Tokenizer System

After receiving the accuracy rating of a claim, it is corrected. This is done by masking each word in a claim and replacing it with the top predicted word from the NLP models. Even if some parts of a claim are correct, they will be replaced. However, this does not matter because if they are truly accurate, the models should be able to predict them, thus, not changing the accurate parts of the original claim at all. Once a claim is corrected, it is passed to the Wikipedia Tokenizer System, where the corrected claim is broken into tokens. These tokens are then searched for on Wikipedia and the lines in the articles for each token are stored. The stored information is then used to create a knowledge graph for the entire claim.
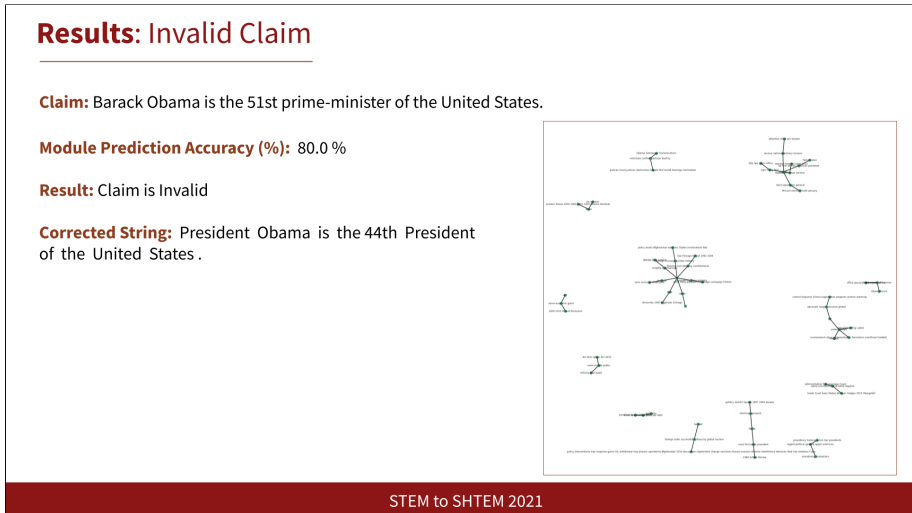
Fig. 5 Actual results from the AI Fact Checking System
Example of an invalid claim

3.4 Fact Check Mobile App

The Fact Check mobile app is an extension of our AI Fact Checking and Claim Correcting system. The name, "Fact Check," has been registered on the Apple App Store and the app currently has a working prototype.
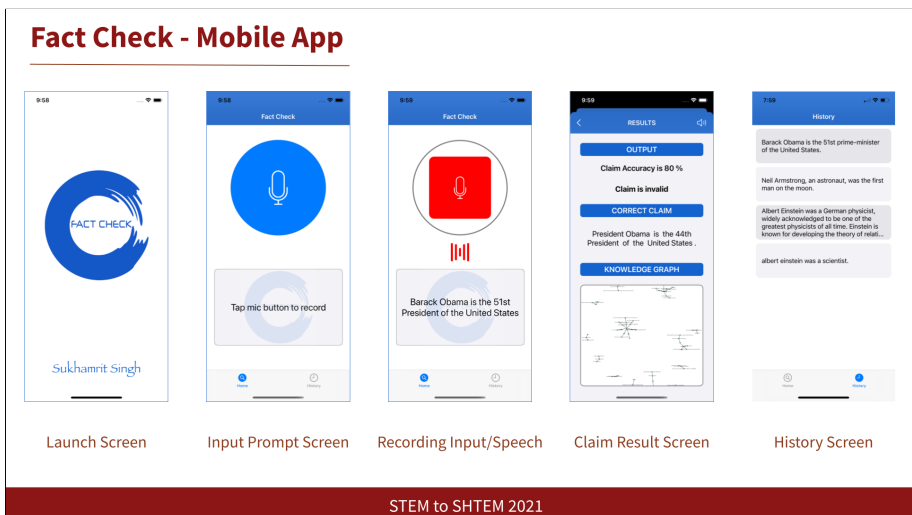

Fig. 6 Screenshots of each screen from the Fact Check Mobile App

When the app is first launched, the Launch Screen is displayed to the user. From there, the user can switch between the Home Screen and the History

Screen by switching tabs on the tab bar located at the bottom of the app (see Figure 6). When on the Home Screen, tapping the microphone button will allow the user to record a claim to be fed into the AI Fact Checking and Claim Correcting System. The app uses the SiriKit NLP from Apple to convert a user's speech into text. Once the speech is converted to text, the claim is passed to our back end hosting the AI Fact Checking and Claim Correcting System. The claim is passed into the system, and the back end then returns the accuracy results, the corrected claim, and an image of a generated knowledge graph (see Figure 7). These results are then shown to the user in the Results Screen. The knowledge graph can be zoomed into to see more details and tapping the speaker icon on the Results Screen plays back the results to the user. The user can also access the History Screen by switching to the History Tab. The History Screen contains all the previously run claims by the app. The results of these previous claims can also be accessed by tapping on one of the claims (see Figure 6).
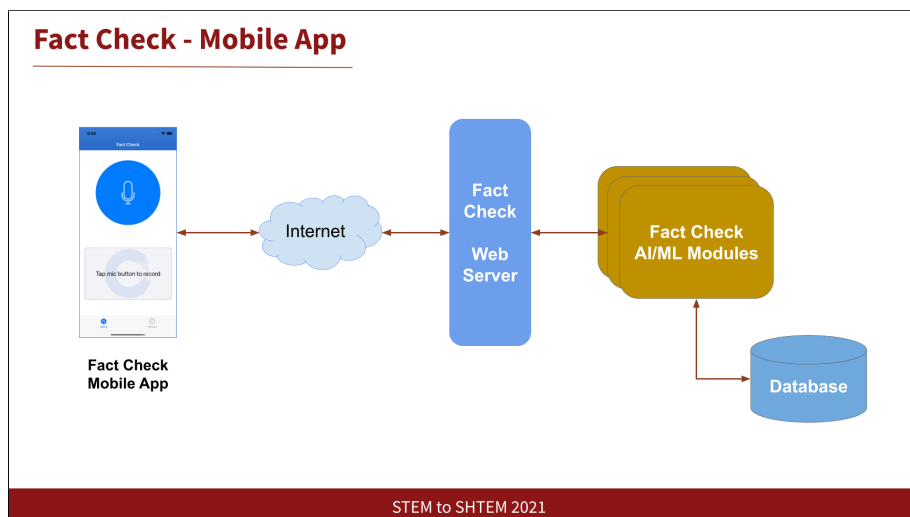

Fig. 7 Fact Check System Block Diagram

4 Discussion

When testing the results of our AI Fact Checking and Claim Correcting System, we compared the results of both DistilBERT and RoBERTa AI models. We documented these results and compared them with data from an online database such as Wikipedia. It was then our group found that both AI models only returned partly accurate predictions. For example, the DistilBERT model was better for predicting numerical components of a claim such as years or size, whereas the RoBERTa model was better at predicting general knowledge such as who the first president of the United States was.

Knowing that both these aspects of a claim are equally as important, our group decided to use both models. We updated our system to follow certain conditions when using these AI models: when the current word of a claim is a number, use the DistilBERT model, and when the current word has a part of speech, such as a noun or verb, use the RoBERTa model. After implementing these conditions, we again tested, documented, and compared the results of our AI Fact Checking and Claim Correcting System with information from Wikipedia. We found that our new results were significantly closer to the data from Wikipedia, indicating our system is providing accurate results.

In the context of the current NLP and fact checking landscape, having to use two types of AI models indicates that current NLP systems are still in need of much work. The use of two models takes up valuable resources such as memory and space, resulting in longer execution times. Current and upcoming models need to be pre-trained with larger amounts of data so that they can provide predictions with greater accuracy and confidence and perform tasks that do not require additional NLP models.

5 Conclusion

The Fact Check System is being built to combat the widespread issue of misinformation. To achieve this, we created an AI Fact Checking and Claim Correcting System. This system takes a claim from the user and scores how accurate it is using the DistilBERT and RoBERTa AI models. It then corrects the claim using the predictions from these AI models and provides a knowledge graph using the corrected claim.

Though there are already AI fact checking models that exist, they do not provide an end-to-end solution for people to use. With the Fact Check application we are building, we are able to provide a solution that users can interact with using a mobile app. The mobile app will not only provide an accuracy rating for a claim, but will also provide a corrected claim if applicable. The Fact Check mobile app is unique in the aspects that it is the first of its kind and it gives users the power of fact checking through something as portable and compact as their phone.

6 Future Directions

We plan to improve our AI Fact Checking and Claim Correcting System by adding a weighting system when it comes to determining a score of accuracy for an inputted claim. Some parts of a claim, such as nouns and verbs, are worth more if they are accurate than other words, such as articles. By adding this weighting system, the accuracy score of a claim is more representative of how accurate the real content of the claim is, overall making the accuracy scores more valid.

Another goal for the future is to add extra features to the Fact Check mobile app to give the user more freedom when inputting a claim. Currently, the Fact Check app only has one method of receiving a claim from the user, which is as a spoken query. However, this method is not always the most optimal. Thus, we plan to add an extra option where the speaker can simply type in their claim to pass to the back end hosting our AI Fact Checking and Sentence Correcting System. We also aim to not only develop and release a mobile app for the Apple App Store but also for the Google Play Store. Creating an app for multiple platforms will allow for a wider audience to have the ability to fact check a variety of information and acts as a solution to the spreading of misinformation.

The source code for both the Fact Check mobile app and the AI Fact Checking and Sentence Correcting System can be found at: https://github.com/Sukhamrit-Singh/Fact-Check

7 Acknowledgments

We want to express our utmost gratitude to our mentor Aadit Trivedi for introducing us to the topic of fact checking using AI and guiding us through our journey of creating our AI Fact Checking and Claim Correcting System. Additionally, we want to thank Professor Tsachy Weissman, Cindy Nguyen, and the Stanford Compression Forum for providing our group this internship opportunity. This program is a valuable stepping stone in our education on AI and without it, the creation of our AI Fact Checking and Claim Correcting System would not have been possible.

8 References

[1] Carterart. (2016, March 16). Hand Drawn Circle Shape Set Free Vector. Vecteezy. https://www.vecteezy.com/vector-art/ 108435-hand-drawn-messy-circle-shape-set.

[2] Bouziane, M., Perrin, H., Cluzeau, A., Mardas, J., amp; Sadeq, A. (2020). Team Buster.ai at CheckThat! 2020: Insights And Recommendations To Improve Fact-Checking. DEI - Unipd. http://www.dei.unipd.it/~ferro/ CLEF-WN-Drafts/CLEF2020/paper_134.pdf.

[3] Ding, Y., Guo, B., Liu, Y., Liang, Y., Shen, H., amp; Yu, Z. (2021). MetaDetector: Meta Event Knowledge Transfer for Fake News Detection. arXiv. https://arxiv.org/pdf/2106.11177.pdf.

[4] Edrisian, A. D. (2016, August 9). Building a Speech-to-text app using speech framework in iOS 10. AppCoda. https://www.appcoda.com/ siri-speech-framework/.

[5] Jones, W. (2015, August 17). Text-to-Speech in Swift in 5 lines. Medium. https://medium.com/@WilliamJones/text-to-speech-in-swift-in-5-lines-e6f6c6139086.

[6] Lazarski, E., Al-Khassaweneh, M., amp; Howard, C. (2021). Using NLP for Fact Checking: A Survey. MDPI. https://www.mdpi.com/2411-9660/5/3/42/pdf.

[7] Liddy, E. (2001). Natural Language Processing . Syracuse University. https://surface.syr.edu/cgi/viewcontent.cgi?article=1019&context=cnlp.

[8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., amp; Stoyanov, V. (2019). Roberta-Base · Hugging Face. roberta-base · Hugging Face. https://huggingface.co/roberta-base.

[9] Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barron-Cedeño, A., Papotti, P., Shaar, S., amp; Da San Martino, G. (2021). Automated Fact-Checking for Assisting Human Fact-Checkers. arXiv. https://arxiv.org/pdf/2103.07769.pdf.

[10] Ninjaprox. (2020). Ninjaprox/Nvactivityindicatorview: A collection of awesome loading animations. GitHub. https://github.com/ninjaprox/NVActivityIndicatorView.

[11] Sanh, V., Debut, L., Chaumond, J., amp; Wolf, T. (2021). Distil-BERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv.https://arxiv.org/pdf/1910.01108.pdf.

[12] Suleiman Khan, P. D. (2021, May 18). BERT, RoBERTa, distilbert, XLNET - which one to use? Medium. https://towardsdatascience.com/bert-roberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8.

[13] Thorne, J., amp; Vlachos, A. (2017). An Extensible Framework for Verification of Numerical Claims. ACL Anthology. https://aclanthology.org/E17-3010.pdf.

[14] Thorne, J., amp; Vlachos, A. (2018). Automated Fact Checking: Task formulations, methods and future directions. ACL Anthology. https://aclanthology.org/C18-1283.pdf.

[15] Thorne, J., amp; Vlachos, A. (2021). Evidence-based Factual Error Correction. arXiv. https://arxiv.org/pdf/2012.15788v2.pdf.

[16] Vlachos, A., amp; Riedel, S. (2014). Fact Checking: Task definition and dataset constructionAndreas Vlachos. ACL Anthology. https://aclanthology.org/W14-2508.pdf.