

Exploring Unsupervised Semantic Similarity Methods for Claim Verification in Health Care News Articles

Vishwani Gupta¹, Astrid Viciano², Holger Wormer³, and Najmehsadat Mousavinezhad¹

¹Fraunhofer IAIS, Sankt Augustin, Germany

²Medien-Doktor Gesundheit, Institut für Journalistik, TU Dortmund, Dortmund, Germany

³Lehrstuhl für Wissenschaftsjournalismus, Institut für Journalistik, TU Dortmund, Dortmund, Germany

{vishwani.gupta, Najmehsadat.Mousavinezhad}@iais.fraunhofer.de

astrid.viciano@tu-dortmund.de

holger.wormer@uni-dortmund.de

Abstract

In the 21st century, the proliferation of fake information has emerged as a significant threat to society. Particularly, healthcare medical reporters face challenges when verifying claims related to treatment effects, side effects, and risks mentioned in news articles, relying on scientific publications for accuracy. The accurate communication of scientific information in news articles has long been a crucial concern in the scientific community, as the dissemination of misinformation can have dire consequences in the healthcare domain. This paper delves into the application of unsupervised semantic similarity models to facilitate claim verification for medical reporters, thereby expediting the process. We explore unsupervised multilingual evidence retrieval techniques aimed at reducing the time required to obtain evidence from scientific studies. Instead of employing content classification, we propose an approach that retrieves relevant evidence from scientific publications for claim verification within the healthcare domain. Given a claim and a set of scientific publications, our system generates a list of the most similar paragraphs containing supporting evidence. Furthermore, we evaluate the performance of state-of-the-art unsupervised semantic similarity methods in this task. As the claim and evidence are present in a cross-lingual space, we find that the XML-RoBERTa model exhibits high accuracy in achieving our objective.

among 600 adults in Florida, revealing substantial exposure to COVID-19 vaccine misinformation among participants. Approximately 73% reported encountering misinformation in the past six months. An overview of current fake news research is given by (Kim et al., 2021). Through the convergence of computational and social science research, they delve into the significance and trajectory of enhancing "digital media literacy" in diverse contexts of news generation and consumption.

Detecting misinformation has emerged as a critical challenge due to the rapid dissemination of news and the potentially severe consequences associated with false information. However, only a limited number of approaches have been developed to address the dynamic, versatile, and fast-spreading nature of fake news editorials. This challenge becomes even more pronounced in the healthcare domain, where the availability of training data is scarce, and pre-trained models may not be readily applicable. While supervised models rely on manually annotated training data, an unsupervised evidence retrieval and verification approach proves more suitable for quick response and works effectively with low-resource languages and domains.

The German HealthNewsReview project *medien-doktor.de* at TU Dortmund University evaluates the quality of medical reporting in German-speaking countries. In this paper, we aim to develop a semi-automated tool that will support journalists in their daily work by evaluating the quality of their ongoing reporting and, also, by finding scientific claims in research papers and journalistic articles with a team of highly renowned medical reporters. The medical reporters evaluate the quality of medical reporting in German-speaking countries and assess the quality of print, radio, online, and TV contributions by applying a catalog of criteria in a journalistic peer review process as explained by (Anhäuser et al., 2020). The detailed criteria have

1 Introduction

The rise of misinformation has been greatly amplified by the advent of social media, primarily due to its increased dissemination and influence. One prominent manifestation of this issue is vaccine hesitancy, which has had significant societal repercussions. To illustrate this, a web-based survey (Neely et al., 2022) was conducted in June 2021



Using resting-state fMRI, we characterized a circuitry of interest consisting of three network modules interconnected by the salience network that displays reduced network coupling in preterm compared with full-term newborns. Interestingly, preterm infants exposed to music in the neonatal intensive care units have significantly increased coupling between brain networks previously shown to be decreased in premature infants: the salience network with the superior frontal, auditory, and sensorimotor networks, and the salience network with the thalamus and precuneus networks. Therefore, music exposure leads to functional brain architectures that are more similar to those of full-term newborns, providing evidence for a beneficial effect of music on the preterm brain.



Die Gehirne aller Kinder wurden nach einigen Wochen im Computertomografen untersucht, zum Vergleich auch die von Babys, die nicht zu früh geboren worden waren. In den Gehirnen der Frühchen, die regelmäßig die Musik gehört hatten, waren einzelne Regionen besser miteinander vernetzt, ergab die Auswertung der Bilder. Ihre Gehirne ähnelten damit stärker denen der später und reifer geborenen Babys.

Figure 1: We are interested in measuring the information similarity of statements in the scientific findings and news, shown here with real examples.

been developed following the example of international research projects such as healthnewsreview.org in the USA as discussed by (Schwitzer, 2008). The detailed evaluations are published on the website medien-doktor.de, along with advice on scientific reporting, media analyses, and blog posts on selected topics. Target groups are not only journalists, but also communication officers at research institutions, teachers, and lay citizens interested in improving their media and scientific literacy. In some newsrooms (among others German Press Agency, WDR, ZDF), these criteria of Medien-Doktor have already been taught as a possible standard for early-career reporters. Nevertheless, many non-specialized newsrooms still lack quality standards in science and medical reporting, particularly among regional media. In contrast to large national media with well-established science sections, regional newspapers often lack editors with scientific backgrounds. As analyses of evaluated articles have shown over the past years, the quality of medical reporting in local journalism usually lags behind the standards of national media. Nevertheless, especially in the German media landscape regional media still contribute significantly to opinion-forming and decision-making in wide circles of the population, while at the same

time suffering the most from the loss of advertising income and structural upheaval in the time of changing habits of media usage. We, therefore, propose here the first steps towards quality-assuring tools that will help regional but also other media with their daily health reporting by economizing editorial resources.

As an initial step towards developing a semi-automated tool, we focus on the "positive effects" criteria from the criteria catalog, which assesses how the potential benefits of therapies, tests, products, or procedures are presented. Journalists need to find evidence supporting claims made in scientific publications, a manual and time-consuming process. This presents a major challenge for healthcare reporters, as they rely on scientific publications to verify claims in news articles.

Healthcare news reporting is further complicated by the fact that journalists often need to translate highly technical language into layperson-friendly terms, as they disseminate scholarly information to audiences outside the research community, including the general public and policymakers. The public relies on the media to learn about new scientific findings, and media portrayals of science significantly influence people's trust in science and their subsequent actions. However, there is a risk

of inadvertently spreading misinformation in this process.

In this paper, we leverage recent advancements in Natural Language Processing, specifically the Transformer architecture, to develop a semantic-aware multilingual Transformer-based architecture for unsupervised evidence retrieval in healthcare claims. We propose an evidence retrieval approach instead of treating the issue as a simple classification task, thus aiding journalists by providing a list of supporting evidence and reducing their manual workload.

We present an architecture that assists fact-checking journalists in verifying the veracity of claims by contextually comparing them against evidence found in scientific publications. This paper addresses the following challenges:

- Finding similarity between scientific evidence and paraphrased scientific findings.
- Extracting evidence across different languages in news articles, considering that most scientific journals and evidence sources publish in English while we work with German news articles.

Both these challenges are demonstrated by an example showcased in Fig. 1.

2 Related Work

The state-of-the-art methods for misinformation detection deal with claim verification in news articles and involve supervised methods, e.g., (Luken et al., 2018; Rawat and Kanojia, 2021). A good survey is (Guo et al., 2022). Most authors treat evidence retrieval and claim verification as a single task referred to as factual verification, e.g., (Nie et al., 2018). To overcome the main challenge of supervised approaches, i.e., the time and labor-intensive construction of reliably annotated datasets to train supervised models, some groups explore the potential of unsupervised models for misinformation detection, e.g., (Yang et al., 2019; Li et al., 2014). Independent of the modeling approach, the reliability of a source plays an important role in evidence retrieval and the verification process. Some work has been done to explicitly compute the reliability of a source, e.g., (Yan et al., 2022). In this section, we will briefly present representative results for each category.

The authors of (Nie et al., 2018) present a connected system consisting of three homogeneous

neural semantic matching models that conduct document retrieval, sentence selection, and claim verification jointly for fact extraction and verification.

In (Luken et al., 2018), the authors break down the process into three modules: potentially relevant documents are gathered based on key phrases in the claim, then sentences relevant to the claim are extracted as evidence from these documents, and finally, the classifier discards any evidence deemed irrelevant and uses the remaining to classify the claim’s veracity. An approach in which the evidence is gathered automatically for each claim is proposed in (Rawat and Kanojia, 2021). The approach extracts supporting evidence from the web articles and then selects appropriate text to be treated as evidence sets. A pre-trained model is used to summarize these evidence sets and then these extracted summaries are used as supporting evidence to aid the classification task. The approach collects evidence and prunes to top-k-related news items based on semantic similarity via BERTScore.

In (Wu et al., 2020) the authors proposed integrating credibility assessment as a part of the fact-checking task. The model first strengthens the interaction between claims and relevant articles to discover key evidence fragments, and then incorporates source features of articles and mitigates the interference of extreme semantics to explore more credible evidence discussing the questionable parts of claims.

In (Li et al., 2014), authors worked on the problem of automatically identifying trustworthy information and sources from multiple conflicting data sources. The authors propose to model the conflict resolution problem on data of heterogeneous types using a general optimization framework called CRH that integrates the truth-finding process on various data types seamlessly. They model the problem using an optimization framework where truths and source reliability are defined as two sets of unknown variables. The objective is to minimize the overall weighted deviation between the truths and the multi-source observations where each source is weighted by its reliability. In (Yin et al., 2008), authors designed a general framework for the Veracity problem and invent an algorithm, called TRUTHFINDER, which utilizes the relationships between websites and their information, i.e., a website is trustworthy if it provides many pieces of true information, and a piece of information is

likely to be true if it is provided by many trustworthy websites. An iterative method is used to infer the trustworthiness of websites and the correctness of information from each other. In (Yang et al., 2019), the authors follow an unsupervised approach by leveraging a Bayesian network model to capture the conditional dependencies among the truths of news, the users’ opinions, and the users’ credibility and proposed an efficient collapsed Gibbs sampling approach to infer the truths of news and the users’ credibility without any labeled data.

In (Yan et al., 2022), authors propose a novel reputation model to quantify the newly defined source reliability, which will be accumulated as the long-term source quality. They propose a reputation-based truth discovery model, where initial weights are assigned based on source reputations. In (Baly et al., 2018), the authors presented a study on predicting the factuality of reporting and bias of news media. The models use a rich set of features derived from the content of the articles from the target news medium, its Wikipedia page, its Twitter account, and information about the web traffic it attracted. In (Mukherjee and Weikum, 2015) the authors analyzed the effect of different factors like language, topics, and perspectives on the credibility rating of articles in a news community. These factors and their mutual interactions are the features of a novel model for jointly capturing the credibility of news articles, the trustworthiness of news sources, and the expertise of users.

Most of the state-of-the-art methods make use of supervised-based models. This will be a challenge when we don’t have annotated data to train large models. In this work, we explored unsupervised based semantic models since in our use case, we have only 20 manually annotated articles. Instead of tackling the problem of claim verification as a classification problem, we propose supporting the journalists with a list of evidence from scientific journals for that given claim. The unsupervised approaches in the literature depend on the characteristics of the news source and the features extracted from the article. These approaches do not consider the semantics and context of the text in the article.

3 Dataset

This paper investigates the correlation between claims made in health news articles and the supporting evidence found in scientific publications. The claims and evidence were manually annotated by

medical reporters due to the labor-intensive nature of this task. Our dataset comprises 20 meticulously annotated articles from prominent German news sources, including Focus Online, Berliner Zeitung, Bild, and Welt. These healthcare news articles encompass a range of topics, such as the positive effects of different treatments/medications, including vaccines for COVID-19, and the relationship between aspirin and the coronavirus.

To substantiate these claims, medical reporters typically refer to scientific publications published in esteemed journals like Nature, PubMed, and Lancet. However, our particular use case presents a multilingual challenge as the news articles are in German, while the scientific studies are in English. The annotated claims consist of a collection of sentences, and correspondingly, the evidence paragraphs in the scientific publications are annotated by the journalists. As part of this ongoing project, we are curating this dataset, which will be made available for future publication.

4 Background

The assessment of text similarity has garnered significant attention from researchers in the fields of natural language processing and information retrieval. This longstanding problem is inherently complex, leading to the development of diverse approaches aimed at capturing a wide range of characteristics. The evaluation of semantic similarity can be categorized into two primary methods: sentence-embedding-based approaches and word-alignment-based approaches.

4.1 Word-Alignment-Based Methods

Alignment-based methods measure the word matching degree for sentence similarity evaluation. WMD is a popular alignment-based method. Its extensions are widely used in text similarity tasks.

4.1.1 Word Mover’s Distance

Earth mover’s distance (EMD), also known as the Wasserstein distance, is a distance measure between two probability distributions. Kusner et al. (Kusner et al., 2015) proposed a version of EMD applicable to language models, the Word mover’s distance (WMD) which evaluates the distance between two documents represented in a continuous space using word embeddings such as the Word2Vec and fastText embeddings. For any two documents A and B , WMD is defined as the minimum cost of transforming document A into docu-

ment B . Each document is represented by the relative frequencies of its words relative to the total number of words of the document, i.e., for the j th word in the document,

$$d_{A,j} = \text{count}(j) / |A| \quad (1)$$

where $|A|$ is the total word count of document A and $\text{count}(j)$ is number of occurrences of the word with vocabulary index j . The j th word is represented by its corresponding word embedding, say $\mathbf{v}_j \in \mathbb{R}^n$. The n -dimensional word embeddings are obtained from a pre-trained model, e.g. Word2Vec or fastText. The distance between two words can easily be measured using Euclidean distance,

$$\delta(i, j) = \|\mathbf{v}_i - \mathbf{v}_j\| \quad (2)$$

Based on this choice, the Word mover's distance is defined to be the solution of the following linear program,

$$\begin{aligned} WMD(A, B) = & \min_{\mathbf{T} \geq 0} \sum_{i=1}^V \sum_{j=1}^V \mathbf{T}_{i,j} \delta(i, j) \\ \text{such that} & \sum_{i=1}^V \mathbf{T}_{i,j} = d_{A,j} \\ \text{and} & \sum_{j=1}^V \mathbf{T}_{i,j} = d_{A,i} \end{aligned} \quad (3)$$

Here, $\mathbf{T} \in \mathbb{R}^{V \times V}$ is a non-negative matrix, where $\mathbf{T}_{i,j}$ denotes how much of word i in document A is assigned to tokens of the word j in document B . Empirically, WMD has reported improved performance on many real-world classification tasks as demonstrated in (Kusner et al., 2015). The WMD has intriguing properties. The distance between two documents can be broken down and represented as the sparse distances between a few individual words. The distance metric is also hyperparameter-free. The most important feature is that it incorporates the semantic information encoded in the word embedding space and is agnostic to arbitrary word embedding models.

4.2 Text Embedding Methods

In such approaches, the aim is to extract a numerical representation of a sentence to encapsulate its meanings. In these methods, we generate embeddings for both claim and the evidence paragraph. The semantic similarity score is calculated using cosine similarity between claim and evidence embeddings.

4.2.1 TF-IDF

The TF-IDF algorithm is a commonly used technique in the extraction of text feature words based on statistical methods. It mainly evaluates the importance of a word-to-text and text sets by word frequency. It is mainly composed of two parts: word frequency and inverse text word frequency. In a document, the term frequency (TF) is the frequency at which a word appears in the text, and the result is usually normalized to prevent it from being biased toward longer text. Inverse Document Frequency (IDF) indicates the importance of a word in a text set.

4.2.2 FastText Based Embeddings

(Bojanowski et al., 2016) proposed an approach that is based on the skip-gram model, where each word is represented as a bag of character n-grams. A vector representation is associated with each character n-gram; words are represented as the sum of these representations. We obtain claim and evidence paragraphs' representations from a pre-trained fastText model which is based on the average of N-gram features.

4.2.3 BERT-Based Embeddings

Bidirectional Encoder Representations from Transformers (Devlin et al., 2019) is one of the most powerful context and word representations. BERT is based on the methodology of transformers and uses an attention mechanism. It employs the bi-directional training of the transformer architecture and applies it to language modeling. Unsupervised objectives, including the masked language model and the next sentence prediction, are incorporated. Word-piece tokenization is performed on the text from both the claim and scientific publication and then used as input to a pre-trained BERT model. The BERT model provides contextual embedding for these word pieces.

- **Sentence-BERT:** (S-BERT) proposed by (Reimers and Gurevych, 2019), is a modification of the pre-trained BERT network that uses siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine similarity.
- **Sci-BERT:** A transformer model proposed by (Beltagy et al., 2019), is trained using masked language modeling on a large corpus of scientific text. It leverages unsupervised pretrain-

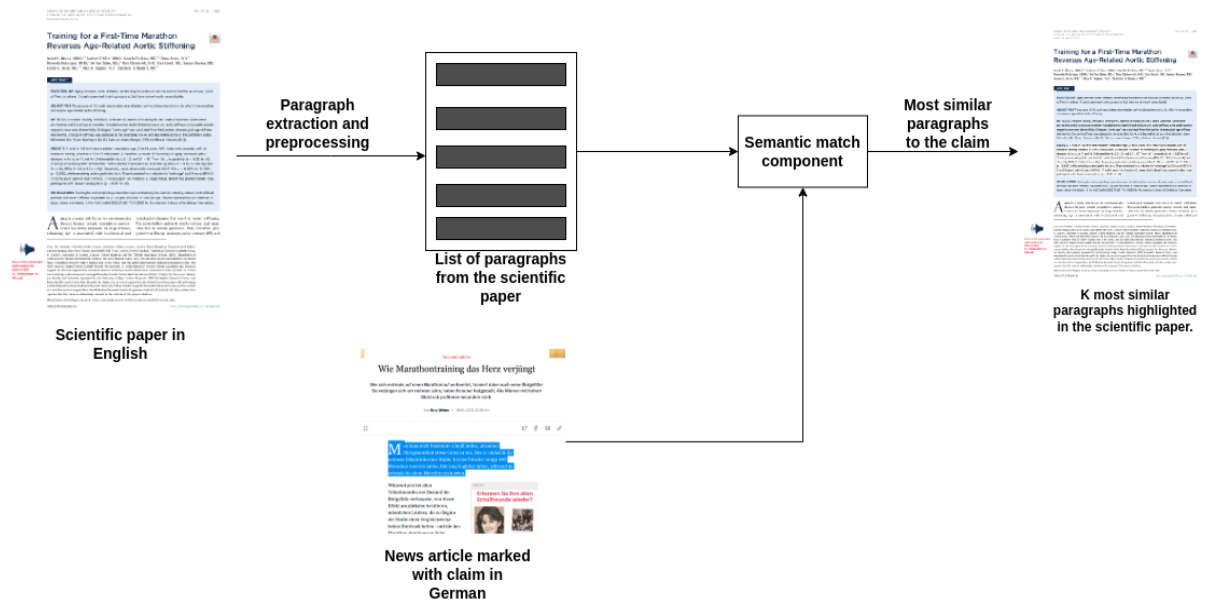


Figure 2: Figure illustrating the pipeline of the semantic match approach for claim verification.

ing on a large multi-domain corpus of scientific publications to improve performance on downstream scientific NLP tasks.

- **XML-RoBERTa:** XLM-R (XLM-RoBERTa, Unsupervised Cross-lingual Representation Learning at Scale) proposed by (Conneau et al., 2020) is a scaled cross-lingual sentence encoder. It is trained on 2.5T of data across 100 languages data filtered from Common Crawl.

5 Proposed Approach

In this section, we describe the architecture of our Semantic Matching component as illustrated in Fig.2. Given a claim in a healthcare news article, we need to find paragraphs in the scientific publication where the evidence for the claim are present. The news article for our use case is in German and the scientific publications are in English, introducing a cross-lingual aspect to the problem. A claim annotated by the medical reporters is a set of sentences in which the positive effects of a medicine or a therapy are explained. As a first step, we extract all the paragraphs from the scientific publication and preprocess the text. Preprocessing in the case of the models trained on English corpus, we translate the claim in English using DeepL translate ¹. The semantic match component takes a claim and the union of the paragraph set from the

¹<https://www.deepl.com/en/translator>

scientific publication as inputs and outputs a subset of paragraphs. Evidence paragraph selection can also be formulated as semantic matching between each paragraph and the claim to select the most plausible evidence set. The selection is done via these steps:

- Calculating the semantic similarity score, s_i , for all the paragraphs in the scientific publication.
- Sorting sentences by their s_i values and adding the top k-paragraphs to the resulting list.

Our proposed pipeline generates a collection of the top k most similar paragraphs, which serve as evidence in the context of fake news detection. Unlike traditional approaches that treat fake news as a classification problem, our pipeline introduces an evidence retrieval approach. This approach effectively assists journalists in locating relevant supporting evidence, thereby reducing the need for manual search efforts. The advantages of our system include: (i) its unsupervised nature, allowing it to adapt to concept drifts without relying on labeled data, and (ii) empowering users with decision-making capabilities while minimizing manual workload.

6 Evaluation and Discussion

In this paper, we employed manual annotation by medical reporters to annotate the evidence for

Model name	German claim translated to English?	Accuracy for k=10	Accuracy for k=5	Accuracy for k=1
<i>tf-idf</i>	Yes	0 %	0%	0%
<i>fastText</i>	Yes	0%	0%	0%
<i>Word Movers distance</i>	Yes	65%	30%	10%
<i>Sentence BERT</i>	Yes	65%	40%	10%
<i>SciBERT</i>	Yes	70%	40%	20%
<i>XML-RoBERTa</i>	No	90%	50%	30%

Table 1: Accuracies score for different semantic similarity-based models for 20 annotated articles.

claims in healthcare news articles, which were predominantly in German. To bridge the language gap, we utilized DeepL, a translation tool, to translate German claims into English. For the experiments, we explored both monolingual and multilingual semantic similarity models. The monolingual models utilized DeepL translations, while the multilingual models, such as XML-RoBERTa, enabled us to handle both English and German texts.

To measure the semantic distance between sentences, we developed a component that searches for semantically similar evidence in scientific studies once a new claim is received. This component employs transformer models to generate representation embeddings for each claim and the paragraphs in the scientific publications. By calculating the similarity distance, we identify the most similar evidence, aiming to provide users with semantically related evidence.

During the evaluation, we considered partial evidence matches within the extracted "k"-nearest neighbors as valid matches. This approach supports journalists in finding additional evidence to supplement partial matches. We evaluated various methods, including word alignment-based approaches and sentence embedding methods, for unsupervised evidence retrieval. The models extracted the "k"-nearest neighbors that exhibited the highest similarity to the given paraphrased claim.

From our results presented in Table 1, XML-RoBERTa demonstrated the best performance in extracting evidence for the given paraphrased claims. Classical semantic similarity approaches using *tf-idf* and *fastText* embeddings did not perform well, as these approaches struggle to capture contextual information effectively. Among the embedding approaches, word movers distance with *fastText* embeddings outperformed cosine similarity measures using *tf-idf* or *fastText*. Word movers distance treats text similarity as a transportation problem,

utilizing word embeddings to determine shared meanings or contextual usage, thereby achieving superior performance compared to cosine similarity models.

In terms of transformer-based models, semantic similarity using XML-RoBERTa embeddings performed the best. Additionally, cross-lingual models outperformed monolingual models, highlighting the benefits of leveraging multilingual capabilities in our approach.

7 Conclusion and Future Work

In this paper, we tackled the challenge of claim verification by employing evidence retrieval techniques from scientific studies. Our approach involved developing a semantic matching method capable of retrieving the most similar evidence from a given scientific report. Through the evaluation of various semantic similarity methods, including text representations and word alignment techniques, we demonstrated the effectiveness of using a multilingual model like XML-RoBERTa to calculate semantic similarity and identify relevant paragraphs containing the evidence. By approaching this as an evidence retrieval rather than a classification problem, our proposed approach aims to support medical reporters and journalists in efficiently locating supporting evidence for paraphrased claims, thereby reducing the need for manual searching.

Moving forward, we will focus on retrieving the most relevant scientific papers from a pool of documents that encompass the supporting evidence for a given claim in healthcare news articles. Additionally, we recognize the need for improvement in the k-nearest neighbors within our models. To achieve this, we plan to target specific sections within scientific publications, such as the results or conclusion section, where there is a higher probability of finding pertinent evidence. These advancements will further enhance the efficacy and precision of our

evidence retrieval approach, paving the way for more accurate claim verification.

Acknowledgments

We would like to thank Tobias Kreutzer and Paul Klammer for their contribution in the project and feedback.

References

- Marcus Anhäuser, Holger Wormer, Astrid Viciano, and Wiebke Rögener. 2020. Ein modulares modell zur qualitätssicherung im medizin-und ernährungsjournalismus. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 64.
- Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '18*, Brussels, Belgium.
- Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. *Scibert: A pretrained language model for scientific text*.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. *Unsupervised cross-lingual representation learning at scale*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *Bert: Pre-training of deep bidirectional transformers for language understanding*.
- Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. 2022. *A Survey on Automated Fact-Checking*. *Transactions of the Association for Computational Linguistics*, 10:178–206.
- Kim, Xiong A, Lee D, and K Han. 2021. *A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions*. *PLoS One*. 2021.
- Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 957–966.
- Qi Li, Yaliang Li, Jing Gao, Bo Zhao, Wei Fan, and Jiawei Han. 2014. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1187–1198.
- Jackson Luken, Nanjiang Jiang, and Marie-Catherine de Marneffe. 2018. *QED: A fact verification system for the FEVER shared task*. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 156–160, Brussels, Belgium. Association for Computational Linguistics.
- Subhabrata Mukherjee and Gerhard Weikum. 2015. Leveraging joint interactions for credibility analysis in news communities. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 353–362.
- Stephen R. Neely, Christina Eldredge, Robin Ersing, and Christa Remington. 2022. *Vaccine hesitancy and exposure to misinformation: a survey analysis*. *Journal of General Internal Medicine*.
- Yixin Nie, Haonan Chen, and Mohit Bansal. 2018. *Combining fact extraction and verification with neural semantic matching networks*. *CoRR*, abs/1811.07039.
- Mrinal Rawat and Diptesh Kanojia. 2021. *Automated evidence collection for fake news detection*. *CoRR*, abs/2112.06507.
- Nils Reimers and Iryna Gurevych. 2019. *Sentence-bert: Sentence embeddings using siamese bert-networks*.
- Gary Schwitzer. 2008. *How do us journalists cover treatments, tests, products, and procedures? an evaluation of 500 stories*. *PLoS medicine vol. 5,5 (2008)*.
- Lianwei Wu, Yuan Rao, Xiong Yang, Wanzhen Wang, and Ambreen Nazir. 2020. Evidence-aware hierarchical interactive attention networks for explainable claim verification. In *IJCAI*.
- Lei Yan, Kan Yang, and Shouyi Yang. 2022. *Reputation-based truth discovery with long-term quality of source in internet of things*. *IEEE Internet of Things Journal*, 9(7):5410–5421.
- Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5644–5651.
- Xiaoxin Yin, Jiawei Han, and S Yu Philip. 2008. Truth discovery with multiple conflicting information providers on the web. *IEEE transactions on knowledge and data engineering*, 20(6):796–808.