Sentiment Impact on Fake News Detection: A Preliminary Study

Luigi Lomasto[1], Raffaele Aurucci[1], Yuri Brandi[1], Lukasz Gajewski[1], Nicola Lettieri[2], Delfina Malandrino[1], and Rocco Zaccagnino[1]

[1] University of Salerno, Fisciano (Italy), Departement of Computer Science
{llomasto,dmalandrino,rzaccagnino}@unisa.it,
{r.aurucci1,y.brandi,l.gajewski}@studenti.unisa.it
[2] National Institute for Public Policy Analysis (INAPP), Rome (Italy), Italy
n.lettieri@inapp.org

Abstract. The *information disorder* phenomenon represents one of the main challenges for the current society, that researchers of a huge variety of scientific areas are trying to solve. To date, the majority of the studies carried out in this context are focused on *Machine Learning*-based Fake News detectors trained on textual data. Despite the numerous attempts available in the literature and the promising results of such models, unfortunately, the expectations are not truly met considering their use in real-world scenarios. The main limitations are directly related to the nature of the phenomenon since the model trained on past events can't understand and classify novel contents and breaking news. In the current work, we study if sentiment discrepancy between news and its associated evidence can determine the possible presence of fake content. In fact checking activities, the evidence are inferred claims or information used to accept or reject shared news [2]. To do that we have conducted a preliminary study in which the outcome is a metric called *Negativity Score*. This metric can be used as a feature in Fake News detection models and automated fact-checking activities. For completeness, we have also provided a framework that uses the proposed metric. The results of preliminary experiments highlight the possibility of exploiting sentiment aspects in addition to more common and well-known approaches.

Keywords: Information Disorder · Fake News · Automated Fact-Checking · Sentiment Analysis · Natural Language Processing

1 Introduction

In the last decade, Social Networks have become the main place to get informed and updated concerning breaking news. Features such as high accessibility, lack of a restrictive check for the shared content, and the malicious motivations of the users make these tools an ideal place for *Fake News* profiling. By definition, Fake News is news that includes deliberate false or misleading statements designed to manipulate or alter facts with specific malicious intent [11]. This phenomenon, which can proliferate across various communication channels including TV stations, newspapers, social media, and the web, poses a significant challenge in contemporary society. Notably, studies have demonstrated that Fake News is seventy percent more likely to be re-shared than trustworthy news, with a dissemination rate six times faster [22]. The repercussions of this issue are profound, affecting public trust in institutions and healthcare [3], influencing electoral outcomes [4], inciting civil wars, manipulating economies, and so on.

To mitigate the spread of Fake News, scientific communities across various disciplines are endeavoring to devise effective countermeasures. Within the journalistic sphere, an increasing amount of resources is being allocated to manual fact-checking activities [20], focusing on the verification of information against trusted sources. However, these manual efforts are resource-intensive, prompting researchers in Computer Science to explore Automated Fact-Checking [19]. This involves leveraging AI and Natural Language Processing (NLP) techniques to study and analyze the syntactic and semantic aspects of news for fake content detection [5]. Additionally, Social Media Analysis has emerged as a relevant approach for studying the propagation patterns of news within social networks, aiming to mitigate its spread [18].

In this study, we present the idea of finding Fake News by looking at news headings with the help of NLP tasks such as Sentiment Analysis (SA). We believe that if two headings talk about the same topic but have opposite sentiments, they can help us spot Fake News. By comparing the sentiment of news heading with the sentiment expressed by the heading of the trusted evidence, we can figure out if it's true or not. The main contributions of this work are as follows:

- A preliminary study that shows the trend of the sentiment in Real and Fake News, to define a metric called *Negativity Score* useful for computing the fake content percentage.
- Preliminary experiments to assess the effectiveness of the proposed methodology; results obtained highlight the possibility of exploiting sentiment aspects in addition to more common and well-known approaches to provide alternative and more complete Automated Fact-Checking solutions.

The source code and the produced data are available on GitHub [3]. The paper is organized as follows: Section 2 focuses on the study of related work. In section 3 we have conducted preliminary studies useful to define the proposed metric. In Section 4 we expose details about the proposed framework. In Section 5 results of our studies are provided. Finally, in Section 6 discuss the conclusion and future works.

2 Related Works

The spread of Fake News represents an interdisciplinary problem that primarily involves Social Sciences, Computer Science, and the journalistic community.

[3] https://github.com/EdgeResearch/Angelia.git

With the emergence of Artificial Intelligence (AI), numerous scientists associated with this research area are endeavoring to address this issue by employing NLP and Deep Learning techniques to detect potentially fake and harmful news, taking into account their contents as elucidated in [13]. The field of Fake News Detection is extensive and offers a variety of solutions, such as *Style-based methods*, *source-based models*, *behavioral-based models*, and *knowledge-based* models.

Style-based methods focus on discerning the writing style of news to identify certain textual patterns characteristic of Fake News. An approach suggested by [6] utilizes syntactic, sentimental, grammatical, and readability features as linguistic attributes to aid Fake News Detection. More recent research activities as in [12] try to outperform, with good results, Fake News Detection by using a bidirectional transformer model like BERT, to consider also the contextual meaning of the text. Another kind of approach employs an Information Propagation Network for Fake News Detection [7] that uses propagation-based concepts, focusing on how news is disseminated online and how the users share content. Furthermore, text clustering-based techniques have been also proposed. The primary challenge with such approaches lies in its susceptibility to noise, though various solutions have been proposed in recent years. For instance, the *pre-processing pipeline* proposed in [9] aims to enhance the efficiency of this technique.

Source-based models assign *credibility* scores or evaluations to sources and use them for Fake News Detection as proposed by [15], [14] and [8].

In the *behavioral-based models*, the social aspects and the behavioral user are considered to counter the Fake News phenomenon. In [10] and [1], the authors address the problem by employing approximate rough set theory to interpret the effects of Information Disorder on social network users, and more in detail about how online communities change their opinion over time.

Lastly, *knowledge-based* models for Fake News Detection frequently employ fact-checking processes to verify the truthfulness of news by comparing the analyzed news with facts. An example of such a model is a framework for Fake News Detection proposed by [21], where pertinent information is extracted from images of FN on social media platforms using Optical Character Recognition systems. Subsequently, this information is used to query Google, and the results are then filtered through a "Verified Sources model". For each filtered result, the heading and content of each article are extracted and inputted into two Convolutional Neural Networks (CNNs), the output of which is finally utilized to determine the veracity of the news.

The proposed solution aims to face the problem using a hybrid approach that combines *Source-based* with *Style-based* strategies. Specifically, as explained above, in this paper we show the possibility of using sentiment analysis as an additional feature to detect Fake News by checking if the heading of the *analyzed* news and the heading of a *trusted evidence*, which talks about the same argument, have opposite polarity.

3 A sentiment-based metric for automated fact-checking

In this section, we present a novel metric useful to detect potential fake content, considering the sentiment polarity expressed. Starting from the assumption that the credibility of news A can be inferred by the comparison between the sentiment score of A's heading and the sentiment score of a trusted evidence's heading, we have conducted preliminary studies to derive insights useful for the metric formulation.

3.1 Hypothesis and preliminary studies

Starting from [21], whose idea is that the credibility of news can be verified by analyzing itself with related evidence, our research is based on the assumption that if the sentiment scores of the news A and evidence B have *opposite polarity* then B contradicts another news A. Considering the limitations of the accessibility of the news content on the web platforms, we have restricted the SA only to the heading of the news. This approach is possible because there are available legal tools (see Section 4.2) to extract such information. Furthermore, the experiments carried out to assess the effectiveness of the proposed approach show promising results, despite the reduced information content used for the SA.

In light of the motivations just exposed, we have built and analyzed a dataset based on Fake News Net [17] containing 444 pairs of Fake News + Evidence headings obtained by manually searching news from trustworthy newspapers. To execute the sentiment analyses on the pairs, we have used a combination (summing their output label) of the following SA tools which are the commonly used solutions for this task.

- **transformers[4]:** Provided by Hugging Face and written in Python, this library stands out for offering a wide range of pre-trained transformer models. We have used *siebert/sentiment-roberta-large-english* [16].
- **TextBlob[5]:** A Python library for text data processing, based on NLTK (Natural Language Toolkit) and Pattern.
- **sentiment[6]:** A JavaScript library specialized in SA.
- **natural[7]:** A JavaScript library, that offers an array of features such as a sentiment analyzer that evaluates the positive or negative nature of a sentence based on a predefined list of words.

We adopted a combination of the libraries just listed by summing the returned labels [-1,0,1] because there is no one best SA library. As shown by the figure 1, in this case, the possible sentiment values range from -4 to 4, considering the number of libraries used.

[4] Hugging Face: transformers docs, https://huggingface.co/docs/transformers/index
[5] TextBlob Docs: https://textblob.readthedocs.io/en/dev/
[6] sentiment repository: https://github.com/thisandagain/sentiment
[7] NaturalNode: natural Docs, https://naturalnode.github.io/natural/

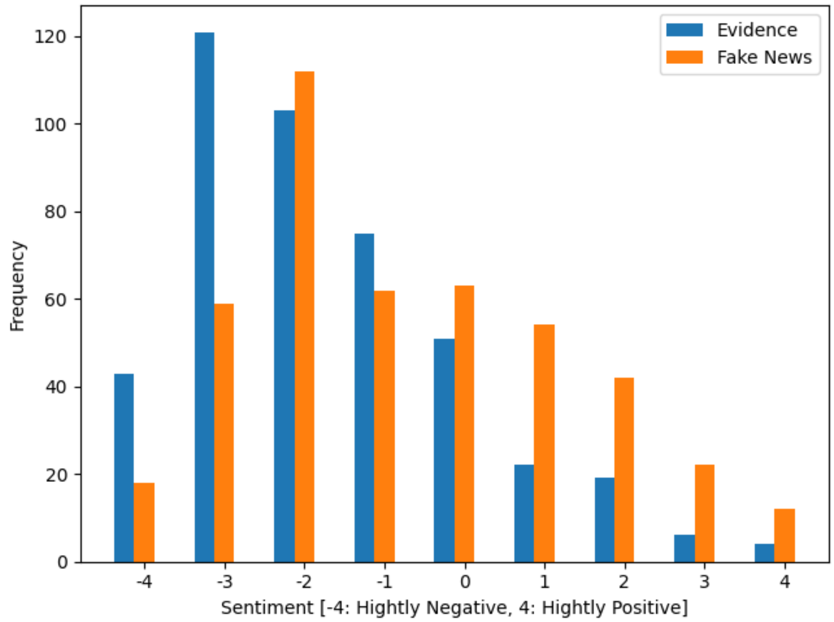

Fig. 1. Sentiment Frequency

It's important to highlight how the sentiment for Evidence, has a high frequency of negative polarity, if compared with the same values of Fake News. These observations led us to analyze the content of the heading for the evidence, at the end of which we noticed that many headings contain *special words* to refute the fake content that influences the sentiment polarity. More in detail, there are 236 evidence for which at least one of the following statements is true:

– Contain the word "*fake*";
– Contain the word "*false*";
– Begin with "*No,* and its variants e.g. 'No.', 'No '".

Continuing with the analysis, we have used the sentiment score in range [-4,4] to establish the sentiment distance between the Evidence/Fake News pair. In this case, the distance values range from 0 to 8. As shown in figure 2, only a limited number of pairs have the same polarity score (∼70). Most of them have mild polarity divergence (e.g. between 1 and 5) and in some cases (e.g. from 6 to 8) the polarities are very distant.

These observations confirm our starting considerations based on the opposition of polarity, and in cases when it's not true, we have understood that there are *special words* used in the evidence to counteract the Fake News.

However, the nature of the problem is complex, and the simple sentiment polarity is not enough to divide sharply the *Real Region* from the *Fake Region*. So we have provided a *Boundary Region*, where Fake and Real News have

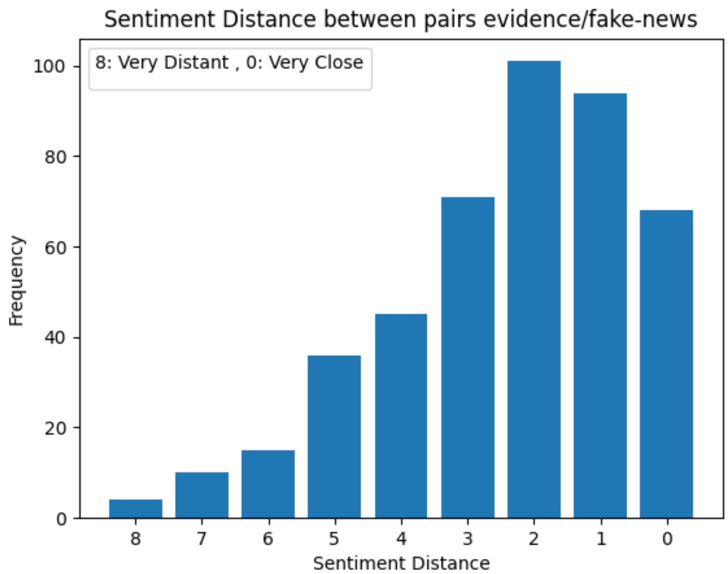

Fig. 2. Sentiment Distance

similar behavior about the sentiment polarity difference considering the related evidence.

Just to understand the next proposed measures, we have to consider that the framework proposed in the section 4, consists of finding, for any news, more trusted evidence. For that, we have to define a metric that uses more evidence for any news to check.

3.2 Metric Definition: The Negativity Score

At this point, we can formalize our *Automatic Fact-Checking Metric*, by introducing the following measures.

- *Successful Check*: check, for each founded evidence, if the sentiments of the news heading and evidence heading are in opposite polarity;
- *Contradictory Score (CS)*: the number of the *Successful Checks*;
- *Neutrality Score (NS)*: the number of evidence with neutral polarity *(=0)*;
- *Evidence Sources (ES)*: number of found evidence to use for news checking.

By using the measures just listed, we define our metric, called **Negativity Score** that takes a high value for fake detected content, as follow:

1. Calculate the percentage of evidence with the same polarity relative to the total number of evidence:

$$PSC = \frac{ES - CS - NS}{ES - NS} \times 100\%$$

2. Finally, calculate the percentage of fake content (*Negativity Score*) as:

$$Negativity\_Score = 100 - PSC$$

In this formula, PSC represents the percentage of the evidence with the same polarity of the news to check without considering NS. By subtracting this percentage from 100%, we obtain the percentage of fake content.

4 Framework

In this section, we propose a framework that aims to use the *Negativity Score* in addition to other machine learning content-based models as in [21], to offer a method for concretely designing solutions based on the proposed metric. To carry out the idea just described, the following *modules* have been defined:

1. *Tokenization* (`TM`): extraction of the main parts *(tokens)* of the analyzed news heading used to perform a web search for *evidence* collection, i.e., news about the same argument of the starting news to check.
2. *Evidence Search* (`ESM`): search of news evidence by using the tokens extracted by `TM`. The evidence is filtered considering the *Trusted Sources List* (`TSL`), a static list of sources which can be considered as *trustworthy* (*trusted sources*).
3. *Polarity Analysis* (`PAM`): compute the sentiment value (positive, neutral, or negative) for the heading of the analyzed news and the set of evidence.
4. *Decision-Making* (`DMM`): establish, by using sentiment scores (computed by `PAM`) and other state-of-art models, if the analyzed news is real or fake.

Starting from a given news heading to check N, the first actions to perform consist of tokenizing it by using *TM*, to obtain the set of relevant tokens T_E and contextually extracting the sentiment value S_N by using *PAM*. The words in T_E are then used by *ESM* to find a set of evidence E_N on the web powered by *TSL*. For each evidence $E_i \in E_N$, the sentiment score S_i is calculated by *PAM*. In the end, the *DMM* component uses the information just retrieved to establish if N is real or fake.

4.1 Tokenization Module (TM)

This module extracts the main tokens of a news heading, which will represent the keywords used by the *ESM* to find evidence. The TM is crucial to performing efficient searches and retrieval of relevant and consistent evidence to evaluate the truthfulness of the startling news. In detail, to implement this module, we have used spaCy [8] library in Python that allows us to use powerful tokenization models like *en_core_web_md* [9].

[8] spaCy Docs, https://spacy.io/api/doc

[9] spaCy: en_core_web_md docs, https://spacy.io/models/en#en_core_web_md

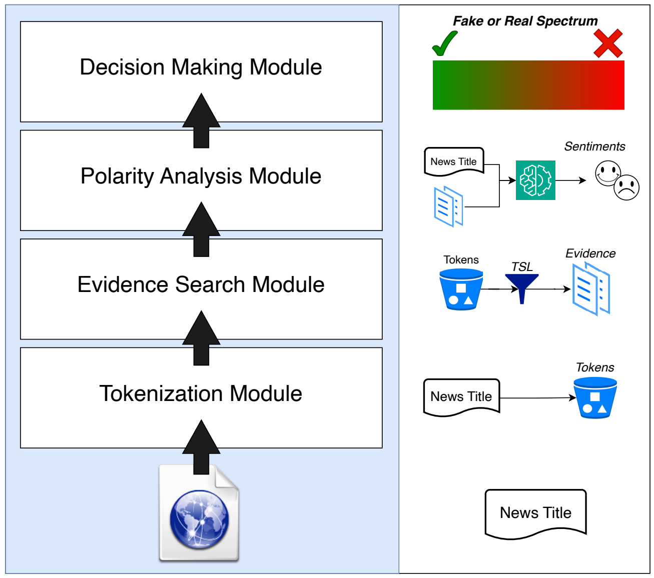

Fig. 3. The System's Architecture is viewed as a layered structure.

4.2 Evidence Search Module (ESM)

The *ESM* makes queries on the web by using the keywords retrieved by the *TM* to find evidence related to the analyzed news. For this activity, we have evaluated two distinct solutions that are *Brave Search API*[10] [11] and Google's *Custom Search JSON API*[12]. Both return a JSON response that contains several information including the *heading*. We opted for *Brave Search API* for its more stringent privacy policies.

Trusted Sources List (TSL) To consider confident evidence, it's essential to establish a list of trusted information sources that will be used to perform the fact-checking task on the news. These sources must be recognized for their impartiality, accuracy, and reliability, to be able to provide an unbiased information evaluation that can be used to confirm or deny the analyzed news. To build the TSL, many trusted sources have been considered as Forbes, MakeUseOf, PureVPN, and others. The outcome of this analysis is the following list:

- nytimes.com
- wsj.com
- bbc.com
- economist.com
- newyorker.com
- ap.org
- reuters.com
- bloomberg.com
- foreignaffairs.com
- theatlantic.com
- politico.com
- c-span.org

---

[10] Brave Search API: https://brave.com/search/api/
[11] https://brave.com/learn/private-vs-personalized-search/
[12] Custom Search API: https://developers.google.com/custom-search/v1/overview

- csmonitor.com
- npr.org
- propublica.org
- eu.usatoday.com
- fair.org
- pewresearch.org
- pbs.org
- cbsnews.com
- theguardian.com
- edition.cnn.com
- nbcnews.com
- forbes.com
- theconversation.com
- upi.com
- journalistsresource.org
- snopes.com
- huffpost.com
- foxnews.com
- dailymail.co.uk
- factcheck.org
- politifact.com
- avclub.com
- bandcamp.com
- deadline.com
- heavy.com
- indiewire.com
- pitchfork.com
- rollingstone.com
- upworthy.com
- variety.com
- vibe.com
- vulture.com
- washingtonpost.com

4.3 Polarity Analysis Module (PAM) and Decision-Making Module (DMM)

The *PAM* module uses the SA libraries introduced in 3 to perform a *Global Sentiment Score*. To normalize the values we map negative polarity (-4,-3,-2, and -1) to -1 and positive polarity (1,2,3, and 4) to 1. Neutral values remain the same (0). The *DMM* module uses the *Negativity Score* metrics in addition to other Fake News detection models, to establish if the news is potentially fake. This module is crucial for the proposed framework's performance.

5 Preliminary experiments

In this section, we present the results obtained by using the *Negativity Score* metric on a dataset composed of 200 fake and 200 Real News with related evidence which are overall 1415 items, since there are about 3 evidence for each Real/Fake news. An example follows in *Table* 1.

Fake News/Real News	Evidence
Bill Gates: "I think Donald Trump will go down to history as one of the greatest presidents, just like Reagan."	No, Bill Gates didn't tweet Donald Trump will be 'one of the greatest presidents'
Beckham divorced his wife. David and Victoria Beckham	David and Victoria Beckham Shut Down Rumors That They're Getting a Divorce
Secret Service Agent Says Obama Is Muslim & Gay In New Tell-All Book	Fake news site falsely claims Obama is a gay Muslim

Table 1. Example of the employed dataset.

5.1 Negativity Score Accuracy

To evaluate the accuracy of the *Negativity Score* metric, we have executed it for each Real/Fake News of the dataset just introduced. As shown in figure 4, the distribution of the Negativity Score has higher overall negativity values in Fake News compared to Real News.

In light of the shown result, we have defined the *Boundary Region* (as introduced in 3) considering *top margin (TM)* = 50 and *bottom margin (BM)* = 30. In this way, all the news with a Negativity Score minor than BM are considered *Real*, and all news with a Negativity Score major than TM are selected as *Fake*. The rest of the news falls in the Boundary Region and is considered as not classifiable.

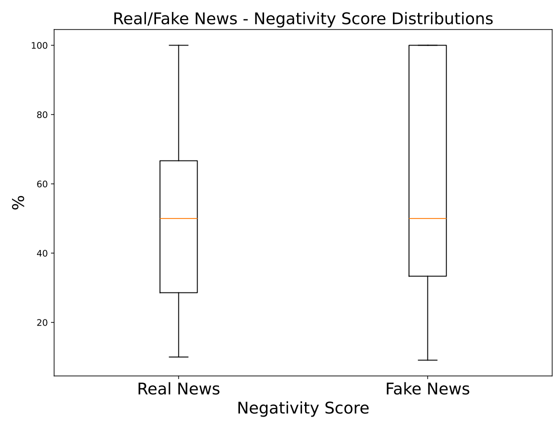

Fig. 4. Box plot for Negativity Score Distributions

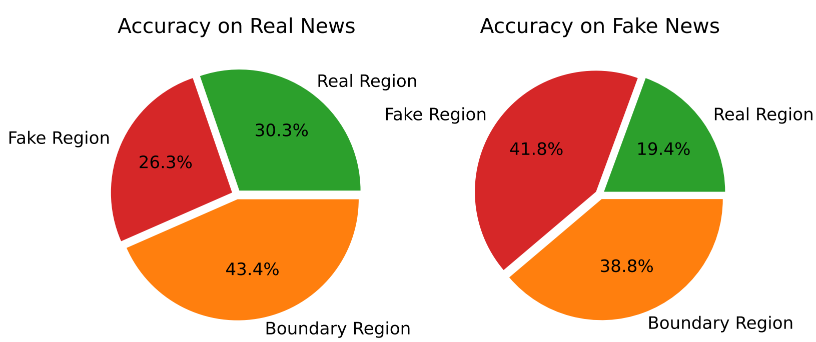

Fig. 5. Negativity Score Accuracy with TM=50 and BM=30

Figure 5 shows the accuracy of the Negativity Score with the defined margins for the Boundary Region, on Fake and Real News. As assumed at the beginning, the results confirm the intuition that, in general, news and evidence with opposite polarity tend to refute each other. The low accuracy on the Real News for the *Real Region* highlights that for real content, the assumptions are not as valid as for the Fake News since the *Boundary Region* is more populated.

6 Conclusions and Future Works

In this study, we introduced a metric to evaluate the sentiment discrepancies between Fake and Real News. Our experiments with specialized datasets and SA tools have shown the possibility of using the proposed metric in more complex fake content detection activities. In this case, the introduction of a *Boundary Region* is essential for fake and real content that shows similar sentiments compared with trusted evidence.

Future work will focus on incorporating multilingual support to improve the overall applicability of the framework. We will delve deeper into the characteristics of the Boundary Region to find discriminant measures that improve the proposed metric. We will also integrate to the machine learning-based solutions proposed by the state of the art, the negativity metric to improve fake news detection.

Acknowledgement
This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

1. Abbruzzese, R., Gaeta, A., Loia, V., Lomasto, L., Orciuoli, F.: Detecting influential news in online communities: An approach based on hexagons of opposition generated by three-way decisions and probabilistic rough sets. Inf. Sci. **578**, 364–377 (2021). https://doi.org/10.1016/J.INS.2021.07.014, `https://doi.org/10.1016/j.ins.2021.07.014`
2. Alhindi, T., Petridis, S., Muresan, S.: Where is your evidence: Improving fact-checking by justification modeling. In: Proceedings of the first workshop on fact extraction and verification (FEVER). pp. 85–90 (2018)
3. Balcaen, P., Buts, C., Bois, C.D., Tkacheva, O.: The effect of disinformation about covid-19 on consumer confidence: Insights from a survey experiment. Journal of Behavioral and Experimental Economics **102**, 101968 (2023). https://doi.org/https://doi.org/10.1016/j.socec.2022.101968, `https://www.sciencedirect.com/science/article/pii/S2214804322001380`
4. Cantarella, M., Fraccaroli, N., Volpe, R.: Does fake news affect voting behaviour? Research Policy **52**(1), 104628 (2023)
5. Capuano, N., Fenza, G., Loia, V., Nota, F.D.: Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review. Neurocomputing **530**, 91–103 (2023). https://doi.org/https://doi.org/10.1016/j.neucom.2023.02.005

6. Choudhary, A., Arora, A.: Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications **169**, 114171 (2021). https://doi.org/https://doi.org/10.1016/j.eswa.2020.114171
7. Cui, B., Ma, K., Li, L., Zhang, W., Ji, K., Chen, Z., Abraham, A.: Intra-graph and inter-graph joint information propagation network with third-order text graph tensor for fake news detection. Applied Intelligence pp. 1–18 (2023)
8. Dias, N., Pennycook, G., Rand, D.G.: Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media (2020)
9. Diaz-Garcia, J.A., Fernandez-Basso, C., Gutiérrez-Batista, K., Ruiz, M.D., Martin-Bautista, M.J.: Improving text clustering using a new technique for selecting trustworthy content in social networks. In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. pp. 275–287. Springer (2022)
10. Gaeta, A., Loia, V., Lomasto, L., Orciuoli, F.: A novel approach based on rough set theory for analyzing information disorder. Appl. Intell. **53**(12), 15993–16014 (2023). https://doi.org/10.1007/S10489-022-04283-9, `https://doi.org/10.1007/s10489-022-04283-9`
11. Gelfert, A.: Fake News: A Definition. Informal Logic **38**, 84–117 (03 2018). https://doi.org/10.22329/il.v38i1.5068
12. Kaliyar, R.K., Goswami, A., Narang, P.: Fakebert: Fake news detection in social media with a bert-based deep learning approach. Multimedia tools and applications **80**(8), 11765–11788 (2021)
13. Mridha, M.F., Keya, A.J., Hamid, M.A., Monowar, M.M., Rahman, M.S..A.: comprehensive review on fake news detection with deep learning. In IEEE **9**, 56151–15617 (2021)
14. Przybyła, P., Borkowski, P., Kaczyński, K.: Countering disinformation by finding reliable sources: a citation-based approach. In: 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2022)
15. Qureshi, K.A., Malick, R.A.S., Sabih, M., Cherifi, H.: Deception detection on social media: A source-based perspective. Knowledge-Based Systems **256**, 109649 (2022)
16. Rozado, D., Hughes, R., Halberstadt, J.: Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models. PLOS ONE **17**, 1–14 (10 2022). https://doi.org/10.1371/journal.pone.0276367
17. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018)
18. Shu, Kai and Wang, Suhang and Lee, Dongwon and Liu, Huan: Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements, pp. 1–19. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-42699-6_1
19. Thorne, J., Vlachos, A.: Automated fact checking: Task formulations, methods and future directions. arXiv preprint arXiv:1806.07687 (2018)
20. Uscinski, J.E., Butler, R.W.: The epistemology of fact checking. Critical Review **25**(2), 162–180 (2013)
21. Vishwakarma, D.K., Meel, P., Yadav, A., Singh, K.: A framework of fake news detection on web platform using ConvNet. Social Network Analysis and Mining **13**(1), 24 (01 2023). https://doi.org/10.1007/s13278-023-01026-7
22. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018). https://doi.org/10.1126/science.aap9559, `https://www.science.org/doi/abs/10.1126/science.aap9559`