

Review

Sentiment Analysis for Fake News Detection

Miguel A. Alonso , David Vilares , Carlos Gómez-Rodríguez  and Jesús Vilares 

Grupo LyS, Departamento de Ciencias da Computación e Tecnoloxías da Información, Universidade da Coruña and CITIC, 15071 A Coruña, Spain; david.vilares@udc.es (D.V.); carlos.gomez@udc.es (C.G.-R.); jesus.vilares@udc.es (J.V.)

* Correspondence: miguel.alonso@udc.es

Abstract: In recent years, we have witnessed a rise in fake news, i.e., provably false pieces of information created with the intention of deception. The dissemination of this type of news poses a serious threat to cohesion and social well-being, since it fosters political polarization and the distrust of people with respect to their leaders. The huge amount of news that is disseminated through social media makes manual verification unfeasible, which has promoted the design and implementation of automatic systems for fake news detection. The creators of fake news use various stylistic tricks to promote the success of their creations, with one of them being to excite the sentiments of the recipients. This has led to sentiment analysis, the part of text analytics in charge of determining the polarity and strength of sentiments expressed in a text, to be used in fake news detection approaches, either as a basis of the system or as a complementary element. In this article, we study the different uses of sentiment analysis in the detection of fake news, with a discussion of the most relevant elements and shortcomings, and the requirements that should be met in the near future, such as multilingualism, explainability, mitigation of biases, or treatment of multimedia elements.

Keywords: sentiment analysis; opinion mining; fake news; social media



check for updates

Citation: Alonso, M.A.; Vilares, D.; Gómez-Rodríguez, C.; Vilares, J. Sentiment Analysis for Fake News Detection. *Electronics* **2021**, *10*, 1348. <https://doi.org/10.3390/electronics10111348>

Academic Editor: Cataldo Musto

Received: 4 May 2021

Accepted: 2 June 2021

Published: 5 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

People are spending more and more time interacting on social media, as the wide adoption of smartphones makes its access available almost anytime and anywhere, which is not the case with traditional media. In addition, they facilitate interaction with friends, families, and even complete strangers through the comment chains, be it through comments, discussions, or simply like and dislike buttons. This has made social media a main channel for the dissemination of news. According to the Pew Research Center's Journalism Project [1], in 2020, 53% of US adults say they obtained news from social media "often" or "sometimes", with 59% of Twitter users and 54% of Facebook users consuming news on the site regularly. Interestingly, 59% of those who obtained news on social media said they expected that news to be largely inaccurate.

Such inaccurate information might result either from a deliberate attempt to deceive or mislead (disinformation) or from an honest mistake (misinformation) [2]. Rumors can fall into either of these two categories, depending on the intent of the source, given that rumors are not necessarily false but may turn out to be true [3]. Unlike rumors, fake news are, by definition, always false and, thus, can be seen as a type of disinformation. Other types of potentially false information that we can find on social media are propaganda, conspiracy theories, hoaxes, biased or one-sided stories, clickbait, and satire news, contributing to information pollution [4]. False information can be propagated by bots, criminal/terrorist organizations, activist or political organizations, governments, hidden paid posters, state-sponsored trolls, journalists, useful idiots, conspiracy theorists, individuals that benefit from false information, and trolls [5]. The motivation behind these actors can be to hurt or disrepute, to obtain financial gain by increasing site views, to manipulate public opinion, to create disorder and confusion, to promote ideological biases, or even as individual entertainment [6].

Sentiment Analysis (SA) is the branch of Natural Language Processing (NLP) in charge of the design and implementation of models, methods, and techniques to determine whether a text deals with objective or subjective information and, in the latter case, to determine if such information is expressed in a positive, neutral, or negative way as well as if it is expressed in a strong or weak way. Since a large part of the subjective content expressed by users on social networks is about opinions (on review sites, forums, message boards, chats, etc.), SA is also known as Opinion Mining (OM).

The expression of sentiment plays an important role in fake news. Social media users tend to comment on posts when there is content that they consider arousing but on which they feel less in control. Conversely, users tend to share a post when they feel more in control [7]. By combining various sentiment variables, Dickerson et al. [8] showed that sentiment-related behavior was sufficient for distinguishing human accounts and social bot accounts. In order to increase the spread of news, headlines should stimulate the reader's curiosity and engage them emotionally. It is not by chance that the spread of fake news is often associated with the presence of clickbait, where the use of the configurations of emotional valence or polarity (positive and negative) and arousal (strong and weak) are knowingly used by publishers to misdirect readers [9] given that a relevant portion of the fake news audience does not read beyond the headlines [10]. Consequently, SA provides crucial information on the content of a news article to determine whether it is trustworthy or should be considered as fake news.

In this article, we discuss the different approaches that have been used to incorporate sentiment in the process of detecting fake news. After providing a description of the related work in Section 2, we define in Section 3 what is meant by fake news and the implication of its dissemination in today's society. We then focus on the process of detecting fake news in Section 4, where we also show a complete list of the data sets that have been built to assist the construction and evaluation of this type of systems as well as the competitive evaluation campaigns that have emerged in recent years; then, we end the section with a description of the measures used to determine the performance of a fake news detection system. In Section 5, we describe SA as a tool that has been used effectively in the broad field of text analytics. Section 6 is dedicated to systems that apply SA in the detection of fake news, considering both those in which SA is the basis of the system and those in which the results obtained by an SA system are used as a feature in a machine learning system. We continue in Section 7 with a discussion of the most relevant elements of such systems, showing in tabular form their main components and proposing the challenges faced in the near future. We finish with the conclusions in Section 8.

2. Related Work

There have been a number of articles reviewing the state-of-the-art in fake news detection in a given moment, but none of them had sentiment analysis or the use of sentiment information as their main focus, as in this case.

Conroy et al. [11] surveyed two major categories of methods for finding fake news. The first one corresponded to linguistic approaches in which the content of deceptive messages is extracted and analyzed to associate language patterns (usage of words, n-grams and syntactic constructions, semantic similarity, and rhetoric relations between linguistic elements) with deception. The second one corresponded to network approaches in which network information, such as message metadata or structured knowledge network queries, can be harnessed to provide aggregate deception measures. In the specific case of the detection of fake reviews, SA was considered a useful technique not specifically to detect fake texts but to detect fake negative reviewers as they overproduced negative emotional terms when compared to truthful reviews as a result of exaggerations of the sentiment they were trying to convey.

Shu et al. [12] described the psychological and social foundations of fake news in traditional and social media and surveyed the features and models used by detection techniques designed to address this phenomenon, considering both news content features and models

as well as social context features and models, which can be based on posts, individual users, or user networks. They considered that SA should play a role in determining post-based features, as people express their emotions or opinions towards fake news through social media posts, such as skeptical opinions or sensational reactions. Years later, Shu et al. [13] revisited the subject by exploring weak social supervision for fake news detection and they stated that user comments that are related to the content of original news pieces are helpful to detect fake news and to explain prediction results. They also considered that machine-generated text created by successful deep generative models can be a new type of fake news that is fluent, readable, and catchy. With respect to sentiment analysis, they still considered sentiment among the features that can be extracted from text for fake news detection given that conflicting sentiments among news spreaders may indicate a high probability of fake news. In [14], Shu and Liu reviewed representative fake news detection methods in a principled way and illustrated challenging issues of fake news detection on social media from a data mining perspective.

Hussein [15] classified 41 articles on sentiment analysis according to the challenge they addressed. They found that eight articles addressed negation, seven dealt with domain dependence, six were dedicated to spam and fake detection, two addressed world knowledge, eight dealt with NLP overheads (sarcasm was included in this challenge), three worked on feature extraction, three studied bipolar words (words in which polarity depends on the context in which they are used [16]), and four dealt with huge lexicons. As can be seen, fake detection is one of the main challenges, although most of the articles analyzed in [15] did not deal with fake news but rather with the detection of fake websites or fake reviews.

Thorne and Vlachos [17] reviewed fact-checking in journalism and listed the resources and methods available to automate such a task as well as the related tasks that could benefit from them. Elhadad et al. [18] differentiated fake news from other forms of disseminating disinformation, misinformation, and malinformation, such as hoaxes, propaganda, satire/parody, rumors, clickbait, and junk news. They added malinformation to the classical categories of disinformation and misinformation. Malinformation was defined as the sharing of genuine information with the intent to cause harm. However, fabricated and junk news, which cannot be considered to contain genuine information, were considered as a possible malinformation realization, which seems contradictory. Sentiment analysis was not mentioned in either [17] or [18].

Bondielli and Marcelloni [19] described the features that have been considered in fake news and rumour detection approaches, provided an analysis on the various techniques used to perform these tasks, and highlighted how the collection of relevant data for performing them is problematic. They considered that the information provided by sentiment analysis techniques can be used to obtain one of the most relevant semantic features of fake news texts.

Sharma et al. [6] addressed fake news detection and mitigation techniques that focus on computational methods to tackle these tasks, compiled a list of available data sets around fake news detection, and proposed a list of challenges and open issues. They found that sentiment analysis was a useful cue for fake news detection, as positive sentiment words tended to be exaggerated in positive fake reviews compared to their true counterparts while replies to fake news on social media tended towards negative sentiment.

Da Silva et al. [20] investigated machine learning approaches and techniques to detect fake news, finding that the preferred methods involved neural networks composed of classical classification algorithms that heavily focus on lexical analysis of the entries as main features for prediction. Sentiment analysis was often used as a content feature in the form of words belonging to sentiment lexicons or as the result of a machine learning-based sentiment analysis system. Klyuev [21] also discussed different approaches to combat fake news and the importance of determining text features by means of natural language processing methods in order to create a profile of the text document. Although he showed

the importance of using dictionaries that contain, among other information, the sentiment polarity of words, he did not explicitly mention the task of sentiment analysis.

Collins et al. [22] described various human-based and machine-based fake news detection models. On the human side, professional fact-checkers are experts in various disciplines who are capable of verifying the veracity of certain news items and decide whether such information is fake or authentic. Human reviewers are overwhelmed by the continued rise in fake news, so the “wisdom of the crowds” can be drawn on through crowdsourcing, based on the premise that no matter how smart someone is, the collective effort of individuals or groups supersedes any single individual’s intellectual capacity. In practice, crowdsourcing is mainly used for the annotation of data sets that are then used for the training of machine learning models rather than for the detection of fake news itself [23]. On the machine side, automated fake news detection techniques use machine learning in conjunction with natural language processing techniques, including sentiment analysis. Hybrid techniques were also considered, in particular an expert-crowdsource approach that combined two methods of manual fact-checking and a human–machine approach that combined machine learning algorithms and the collective effort of humans.

Meel and Vishwakarma [4] surveyed how the contents on the web are contaminated intentionally or sometimes unintentionally by fake reviews, fake news, and satire, among other sources of information pollution. For this purpose, they studied the false information ecosystem, from the categorization of false information and the motivations to spread it to the social impact and user perception. They also discussed the current state of fact-checking, including source detection, propagation dynamics, methods of detection, and methods for containment and intervention. They considered sentiment analysis as one of the base sources of information needed to detect false information.

Oshikaw et al. [24] studied the technical challenges in fake news detection and how researchers define different tasks and formulate machine learning solutions to tackle this task, focusing on how fake news detection was aligned with existing natural language processing tasks.

Zhang and Ghorbani [25] characterized the negative impact of online fake news and studied detection methods for this type of information, finding that many of them rely on identifying features of the users, content, and context that indicate misinformation. They stated that accurate fake news detection is challenging due to the dynamic nature of social media as well as the complexity and diversity of online communication data and that the limited availability of high-quality training data is a big issue for training supervised learning models. They defined each piece of news as consisting of physical news content and non-physical news content, where physical contents are the carriers and formats of the news and where non-physical contents are the opinions, emotions, attitudes, and sentiments that the news creators want to express. As a result, they considered that sentiment analysis is a useful method to illustrate the emotions, attitudes, and opinions that are conveyed by online social media and that sentiment-related factors are key attributes for suspicious account identification.

Zhou and Zafarini [23] surveyed fake news detection from the perspective of knowledge-based methods that detect fake news by verifying if the knowledge within the news text is consistent with facts, style-based methods that are concerned with how fake news is written, propagation-based methods that detect fake news based on how it spreads online, and source-based methods that detect fake news by analyzing the credibility of news sources. They considered sentiment as an important semantic-level feature of news content. They also stated that the implementation of efficient and explainable fake news detection systems needs collaborative efforts involving experts in computer and information sciences, social sciences, political science, and journalism.

De Souza et al. [26] reviewed the different types of features related to fake news detection methods and data sets, and they considered that SA was a useful feature to quickly verify the accuracy of information on social media. Finally, Antonakaki et al. [27] recently presented a survey on current research topics in Twitter, determining that sentiment

analysis was one of the four main branches of research involving Twitter and that one of the major threats for this social network is the dissemination of fake news through it. However, they analyzed both topics separately, without making a connection between the two that shows the usefulness of sentiment analysis in detecting fake news.

3. Fake News

One of the most commonly accepted definitions by the research community is that “Fake news is a news article that is intentionally and verifiably false” [12,28]. Thus, we can identify three key aspects of fake news: its form as a news article, its deceptive intent, and the verifiability of its content as completely or partially false [19]. Wardle [29] deconstructed fake news into seven categories:

- false connection, where headlines, visuals, or captions do not support the content;
- false context, corresponding to genuine content shared with false contextual information;
- manipulated content, i.e., genuine information manipulated to deceive;
- misleading content, which involves misleading use of information to frame an issue or individual;
- imposter content, where genuine sources are impersonated;
- fabricated content, 100% false, designed to deceive and harm; and
- satire/parody, with potential to fool but no intention to cause harm. Given the non-harmful nature of these news and because they are easily identifiable as parodic [30], this type of news is not usually considered for fake news detection, although satire can be used as an excuse to avoid the accusation of spreading false news [31].

Fake news is spreading across all spheres: science [32–34]; art [35,36]; finance [37–39]; marketing [40–42]; politics [43–45]; security [46–48]; defense [49,50]; civil protection [51–53]; and health [54–57], including recent sensitive topics such as vaccines [58,59] and COVID-19 [59–62]. “Fake news” was even named the word of the decade for 2011–2020 by the Macquarie dictionary in 2021 (<https://www.macquariedictionary.com.au/blog/article/780/>, accessed on 3 June 2021) and Word of the Year 2017 by both the American Dialect Society and Collins English Dictionary (https://en.wikipedia.org/wiki/Word_of_the_year, accessed on 3 June 2021). Other related terms also appeared in this classification, for example, “post-truth” was considered Word of the Year 2016 by the Oxford English Dictionary (<https://languages.oup.com/word-of-the-year/2016/>, accessed on 3 June 2021).

The spread of the fake news phenomenon poses a serious negative impact on individuals and society [12], as it can break the authenticity balance of the news ecosystem, it intentionally persuades consumers to accept biased or false beliefs, and it changes the way people interpret and respond to real news. The issue is very serious when considering that fake news spread far more rapidly on social media than real news, especially political news [63].

Bakir and McStay [43] examined the 2016 US presidential election campaign (Trump vs. Clinton) to try to identify problems with, causes of, and solutions to the fake news phenomenon. Fake news were defined as either wholly false news or news containing deliberately misleading elements incorporated within its content or context. They argued that the fake news phenomenon was a logical outcome of five features of the digital media ecology: (1) the financial decline in legacy news, (2) the news cycle’s increasing immediacy, (3) the rapid circulation of misinformation and disinformation via user-generated content and propagandists, (4) the increasingly emotionalized nature of online discourse, and (5) the growing number of people financially capitalizing on algorithms used by social media platforms and internet search engines. Related to (4), they considered sentiment analysis to be an element of great relevance to fake news detection, and they also suggested that the potential to manipulate public sentiment via “empathically optimized” automated fake news was a near-horizon problem. This prediction is particularly worrisome considering that the human ability to detect deception is not good, to the point that it has been shown to be worse than that achieved by machine learning systems. According to [64], human

judgments achieve 50–63 percent success rates, depending on what is considered deceptive, while machine learning algorithms reached 65 percent accuracy at the time of the study.

The interest of the research community in disinformation and misinformation has increased in recent years, with a steady growth in the number of publications on rumors since 2006 [19]. In the case of fake news, there were few publications before 2016, but a rapid growth started in 2017 that has led fake news to become the most important research subject on these issues since 2018, surpassing rumors [19,65].

4. Fake News Detection

The influence of fake news cannot simply be undone by pointing out that the information was incorrect, especially in people with relatively lower cognitive ability [66]. Therefore, it is very important to detect fake news as soon as possible in order to prevent its spread. When the amount of fake news was much less than it is nowadays, the task of checking the veracity of the claims made in news could be carried out by trained journalists. Depending on the complexity of the claim, this process may take from a few minutes to a few days [17]. The enormous growth of social media in the last decade as well as the increasing spread of fake news on the internet makes it impossible to maintain a manual process of checking the news, so it is necessary to resort to automated methods to deal with the veracity of news.

The purpose of automatic fake news detection systems is twofold: first, to reduce human time and effort to detect fake news [24] and, second, to categorize news along a continuum of veracity with an associated measure of certainty, taking into account that veracity is compromised by the occurrence of intentional deceptions [11]. Thus, we can define fake news detection as the process of estimating whether a particular news article of any topic from any domain is intentionally or unintentionally misleading [18]. For Shu et al. [12], fake news is intentionally written to mislead readers to believe false information, which makes it difficult to detect based on news content as it uses diverse linguistic styles while simultaneously mocking true news. Therefore, we need to include auxiliary information, such as user social engagements on social media to detect fake news but this information is big, incomplete, unstructured, and noisy [67].

In recent years, the problem of detecting fake news has been attacked from various angles, leading to multiple different ways of categorizing the different approaches employed for this task [6,12,23,68,69]. One of the most common ways to divide approaches to detecting fake news is to consider the following categories:

- *Knowledge-based, fact-checking approaches* that try to determine whether the claims made in a news story are supported by facts. For this, knowledge bases, including the semantic web and linked open data, are used [70,71]. We could also include in this category fact-checking approaches based on the use of information retrieval techniques to find documents that support a news piece [72,73].
- *Context-based approaches* that try to determine the truthfulness of a news story based on its metadata, such as the credibility of its author and publisher as well as the speed and form of dissemination of the news on social networks [27].
- *Content-based approaches* that try to determine the veracity of a story based on its text, which includes considerations of style (for example, length, variety of words, complexity of vocabulary, and complexity of syntactic constructions) [10,24,68] as well as of the type and strength of sentiments and emotions conveyed by the news, which is the topic covered in this article.

Fake news detection falls within the broader field of misinformation and disinformation detection and mitigation. The approaches to detecting fake news use methods, techniques, and resources that are useful in a variety of other tasks and vice versa. Among those related tasks, we can mention the following ones:

- Stance classification [74,75], the task of determining the opinion behind some text, with the target or topic known in advance (target-specific stance classification) or not

(open stance classification). It is different from fake news detection in that it is not used for assessing veracity but consistency [24].

- Rumor detection [3,19,76]. We can define a rumor as a piece of circulating information for which the truth or falsehood has yet to be verified at the time of spread [3,12]. The resolution of this task involves rumor identification, rumor tracking, stance classification, and veracity classification.
- Truth discovery [77], the task of detecting true facts by resolving conflicts among multi-source noisy information, e.g., databases, the web, crowdsourced data, etc. To solve this task, both the credibility of the source and the truthfulness of the objects of interest need to be taken into account.
- Clickbait detection [9], the task of distinguishing headlines that fit the facts described in a news item from those designed for eye-catching and teasing in online media.
- Opinion spam [78,79] and fake reviews [80,81] detection, the task of filtering out fictitious opinions that have been deliberately written in review sites to sound authentic. People may write this type of fake reviews to promote their products or to defame their competitors' products.
- Bot detection [82]. Social media is now populated by small programs designed to exhibit human-like behavior called social bots [83] that automatically spread posts to give the impression that a given piece of information is highly popular and endorsed by many people. The task of detecting bots is closely related to opinion spam and fake review detection.
- Determining source credibility [84,85]. Credibility is a perceived quality associated with believability, trustworthiness, perceived reliability, expertise, and accuracy. As a result, we could say that credibility is a transversal quality for all tasks that deal with misinformation and disinformation.
- Hate speech detection [86–88]. Hate speech is a broad umbrella term for insulting and offensive user-created content addressed to specific targets to incite violence or hate toward a minority based on specific characteristics of groups, such as ethnic origin, religion, or other. In hate speech, language is used to attack or diminish these groups. These types of messages are based on content of doubtful credibility, including rumors and fake news.

4.1. Resources

Finding the resources and data sets used in fake news research is not an easy task. To find data sets, we followed two processes:

- Papers that discuss experimental results usually provide some reference to the source of the data used in such experiments, which can range from simply indicating the name of the data set to providing a link or, more rarely, a bibliographic reference. In the case of links, a lot of them were broken, so we did our best to obtain the right links.
- A search was performed on Google (<https://www.google.com/>, accessed on 3 June 2021) using “fake news dataset” and “fake news corpus” as query phrases to find additional resources.

For each resource found, an additional search process was carried out to find a bibliographic reference in which the resource is described.

To find shared tasks relevant to fake news detection, we also followed two processes:

- Manual browsing was performed for the most popular evaluation campaigns in NLP, such as SemEval, CLEF, IberLEF, etc.
- In addition, a Google search for “fake news shared task” and “fake news campaign” was performed.

As a result, lists of data sets and shared tasks were compiled. For each resource, a summary of its most relevant characteristics as well as a working link and a bibliographic reference in which it is described in more detail were provided. We consider this compilation itself to be a valuable resource for the scientific community.

4.2. Data Sets

Building data sets to test different detection techniques is a crucial research element in this area. The usefulness of a fake news detection corpus depends on several factors: it should provide both truthful and deceptive instances in text format, the news verification process should be clearly indicated, and there should be homogeneity in length, writing matter, and timeframes [89]. The main issues in constructing such data sets are that the amount of false information is a small fraction of the online content produced every day, even if we restrict our focus on news articles and posts discussing breaking news, and that social media companies nowadays have strict policies concerning the analysis of data produced by their users [19]. Below, we list the data sets that have been built by the scientific community to assess the performance of fake news detection algorithms, techniques, and systems. Since most of the data sets comprise texts in English, we only explicitly specify the language of non-English data sets in the list below. It is also worth noting that we focus only on corpora that are composed mainly of news (or news-related content such as Twitter or Facebook posts), so we do not include those that are based on other types of false information, such as fake reviews [90], rumors [91–93], hoaxes [94], or everyday lies [95].

- Fact Checking corpus (<https://sites.google.com/site/andreasvlachos/resources/>, accessed on 3 June 2021) by Vlachos and Riedel [96]. It contains 106 statements from Channel 4 FactCheck (<https://www.channel4.com/news/factcheck/>, accessed on 3 June 2021) and PolitiFact.com (<https://www.politifact.com/>, accessed on 3 June 2021). Both websites have large archives of fact-checked statements that cover a wide range of issues involving UK and US public life, and they provide detailed verdicts with fine-grained labels that were aligned to a five-point scale of “True”, “Mostly True”, “Half True”, “Mostly False”, and “False”.
- BuzzFeed-Webis Fake News Corpus 2016 (<https://zenodo.org/record/1239675>, accessed on 3 June 2021) [68]. This corpus encompasses 1627 Facebook posts from 9 publishers on 7 workdays close to the US 2016 presidential election. It contains 256 posts from three left-wing publishers, 545 posts from three right-wing ones, and 826 posts from three mainstream publishers. All publishers earned Facebook’s blue checkmark, indicating authenticity and an elevated status within the network. Each post and linked news article was rated “mostly true”, “mixture of true and false”, “mostly false”, or “no factual content” by BuzzFeed journalists.
- BuzzFace (<https://github.com/gstantia/BuzzFace>, accessed on 3 June 2021) [97] is an extension of the previous corpus enriched with 1.6 million Facebook comments and reactions and with additional data from Twitter and Reddit.
- Craig Silverman data sets (<https://github.com/BenjaminDHome/fakenewsdata1>, accessed on 3 June 2021). The data set called Buzzfeed Political News Data contains true news stories and malicious fake news stories from buzzfeednews.com. Other data set, Random Political News Data, contains true news from *The Wall Street Journal*, *The Economist*, BBC, NPR, ABC, CBS, *USA Today*, The Guardian, NBC, and *The Washington Post*; satirical news from The Onion, Huffington Post Satire, The Borowitz Report, The Beaverton, SatireWire, and Faking News; and fake news from Ending The Fed, True Pundit, abcnews.com.co, DC Gazette, Liberty Writers News, Before its News, InfoWars, and Real News Right Now.
- LIAR (https://www.cs.ucsb.edu/~william/data/liar_dataset.zip, accessed on 3 June 2021) [98], with 12.8 K human labeled short statements from PolitiFact.com evaluated for its truthfulness using six labels: “pants-fire”, “false”, “barely-true”, “half-true”, “mostly-true”, and “true”. A rich set of meta-data for the author of each statement is also provided. The statements are sampled from news releases, TV and radio interviews, campaign speeches, TV ads, Twitter messages, debates, and Facebook posts. The most discussed subjects are economy, healthcare, taxes, federal-budget, education, jobs, state-budget, candidate biographies, elections, and immigration.

- Fact Checking data set (<https://hrashkin.github.io/factcheck.html>, accessed on 3 June 2021) [99], a collection of rated statements from PolitiFact.com with additional unreliable news articles from different types of unreliable sources including satire, propaganda, and hoaxes.
- BS Detector data set (<https://www.kaggle.com/mrisdal/fake-news>, accessed on 3 June 2021), sometimes known as Kaggle Fake News data set, contains text and metadata from 244 websites and represents 12,999 posts from 30 days collected by a browser extension called BS detector developed for checking news veracity. This extension searches all links on a given webpage for references to unreliable sources by checking against a manually compiled list of domains. Therefore, the documents in the data set were labeled by software, not by human annotators.
- Fake vs. real news project data set (https://github.com/joolsa/fake_real_news_dataset, accessed on 3 June 2021) was developed by George McIntire and contains 7.8 k news with equal allocation of fake and real news. Half of the corpus comes from political news [100]. Fake news were collected from the BS Detector data set. Real news were collected from media organizations such as *The New York Times*, WSJ, Bloomberg, NPR, and The Guardian during 2015 and 2016.
- CREDBANK data set (<https://github.com/compsocial/CREDBANK-data>, accessed on 3 June 2021) [101] comprises more than 60 million tweets collected over 96 days, grouped into 1049 real-world events. Each tweet was annotated by 30 human annotators from Amazon Mechanical Turk.
- Snopes data set (<https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/impact/web-credibility-analysis/>, accessed on 3 June 2021) [102] contains 1277 true claims and 3579 false claims collected from the fact-checking website Snopes (<https://www.snopes.com/>, accessed on 3 June 2021). A set of 30 reporting articles is provided for each claim, which were obtained by using the text of the claim as query to Google in order to extract the first 30 results.
- Fact-Checking Facebook Politics Pages (<https://github.com/BuzzFeedNews/2016-10-facebook-fact-check>, <https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis>, accessed on 3 June 2021) contains 666 posts from three large right-wing hyperpartisan Facebook pages, 471 from hyperpartisan left-wing pages, as well as 1145 posts from three large mainstream political news pages that were manually rated as “mostly true”, “mixture of true and false”, or “mostly false”. Satirical and opinion-driven posts or posts that lacked a factual claim were rated as “no factual content”.
- FEVER, Fact Extraction and VERification data set (<https://fever.ai/resources.html>, accessed on 3 June 2021) [103], consists of 185,445 claims manually verified against the introductory sections of Wikipedia pages and classified as Supported, Refuted, or NotEnoughInfo. For the first two classes, systems and annotators need to also return the combination of sentences forming the necessary evidence supporting or refuting the claim. The claims were generated by human annotators extracting claims from Wikipedia and mutating them in a variety of ways, some of which were meaning-altering. The verification of each claim was conducted in a separate annotation process by annotators who were aware of the page but not the sentence from which the original claim was extracted. Although this data set does not contain news, we consider it relevant and is included in this list because it shows a way to convert a text with objective facts (similar in some way to a true news item) into a false text (equivalent to a fake news).
- Fake News vs. Satire corpus (<https://github.com/jgolbeck/fakenews>, accessed on 3 June 2021) [31] contains 283 fake news articles and 203 satirical stories focused on American politics, posted between January 2016 and October 2017. The title, a link, and the full text ID are provided for each article. For fake news stories, a rebutting article is also provided that disproves the premise of the original story.

- Fake.Br Corpus (<https://github.com/roneysco/Fake.br-Corpus>, accessed on 3 June 2021) [104,105] composed of 7200 true and fake news written in Brazilian Portuguese: 3600 fake news were manually collected from four Brazilian newspapers while 3600 true news were collected in a semi-automatic way from major news agencies in Brazil, choosing the most similar ones to fake news, with manual verification to guarantee that the fake and true news were in fact subject-related.
- FakeNewsCorpus (<https://github.com/architapathak/FakeNewsCorpus>, accessed on 3 June 2021) by Pathak and Srihari [106] contains 704 fake and questionable articles from Aug.–Nov., 2016, on the topic of the 2016 US election. Each article was manually checked, and two types of labels were assigned to it: a primary label based on the assertions made (“False”, “Partial truth”, or “Opinions”) and a secondary label (“Fake” or “Questionable”).
- NELA-GT-2018 (<https://doi.org/10.7910/DVN/ULHLCB>, accessed 3 on June 2021) [107], a data set that does not annotate the veracity of each news story but rather the truthfulness of the news source. It is made up of 713,534 articles from 194 news and media producers including mainstream, hyper-partisan, and conspiracy sources. It includes the ground truth ratings of the sources from eight independent assessment sites (NewsGuard, Pew Research Center, Wikipedia, OpenSources, Media Bias/Fact Check MBFC, AllSides, BuzzFeed News, and Politifact) covering multiple dimensions of veracity, including reliability, bias, transparency, adherence to journalistic standards, and consumer trust.
- NELA-GT-2019 (<https://doi.org/10.7910/DVN/O7FWPO>, accessed on 3 June 2021) [108] is an updated version of NELA-GT-2018 with 1.12 M news articles from 260 sources. One major change is the removal of NewsGuard labels.
- FA-KES (<https://zenodo.org/record/2607278#.YEpyDGhKhPY>, accessed on 3 June 2021) [109], a fake news data set around the Syrian war. The data set consists of a set of 804 news articles written in English from several media outlets representing mobilization press, loyalist press, and diverse print media annotated as fake or credible, taking the information obtained from the Syrian Violations Documentation Center (VDC) as ground truth. For each news, relevant information (e.g., date, location, and number of casualties) was extracted manually through the crowdsourcing platform CrowdFlower and then matched against the VDC database to deduce whether an article is fake.
- The Spanish Fake News Corpus (<https://github.com/jposadas/FakeNewsCorpusSpanish>, accessed on 3 June 2021) [110] contains 971 news written in Spanish compiled from January to July of 2018 from sources from Spain, Mexico, Argentina, Colombia, the USA, and the United Kingdom including newspapers, media companies, websites dedicated to validating fake news, and websites designated by different journalists as sites that regularly publish fake news. Each article was manually labeled as true if there was evidence that it was published on reliable sites (established newspaper websites or renowned journalist websites). Conversely, an article was labeled as fake if it was contradicted by any article from reliable sites or from websites specializing in the detection of deceptive content or if no other evidence was found about the news besides the source. This corpus was used as training data set at the FakeDeS 2021 shared task.
- FakeNewsNet (<https://github.com/KaiDMML/FakeNewsNet>, accessed on 3 June 2021) [111] contains news from the fact-checking websites PolitiFact.com and GossipCop (<https://www.gossipcop.com/>, accessed on 3 June 2021). In addition to text, it also contains information on user engagement related to the fake and real news pieces in the form of social media posts and their replies, likes, and reposts. Timestamps and locations explicitly listed in user profiles are also provided.
- FakeHealth (<https://github.com/EnyanDai/FakeHealth>, accessed on 3 June 2021) [56] consists of two data sets: HealthStory with news stories reported by news media (e.g., Reuters Health) and HealthRelease with news releases from universities, research

centers, and companies. Each data set includes news contents, news reviews covering explanations regarding ten health news evaluation criteria, social engagements from Twitter, and user networks. The news stories and releases in the data sets have been evaluated by experts on a set of reliability criteria. For the news contents, text is provided along with the source publishers, image links, and other side information. News social engagements include 500 k tweets, 29 k replies, 14 k retweets, and 27 k user profiles with timelines and friend lists.

- HWB (<https://dcs.uoc.ac.in/cida/resources/hwb.html>, accessed on 3 June 2021) [55], a Health and Well-Being fake news data set composed of 500 legitimate news on health from reputable sources (CNN, *The New York Times*, and *The New Indian Express*, among others), manually double-checked for truthfulness and 500 fake news on health from well-reported misinformation websites (Before It's News, Nephef, and Mad World News, among others), manually verified for misinformation presence.
- Weibo-20 [112] (<https://github.com/RMSnow/WWW2021>, accessed on 3 June 2021), a Chinese data set with 3161 instances of fake news (1355 are fake news pieces from the Weibo-16 rumor data set [93], and the rest are news pieces judged as misinformation officially by the Weibo Community Management Center) with their 1,132,298 comments on Weibo and 3201 real news (2351 were real news pieces from Weibo-16, and 850 are new real news) with their 851,142 comments on Weibo.

4.3. Shared Tasks

In recent decades, the text analytics research field has been characterized by the development of collaborative resources that materialized mainly through the organization of “shared tasks”. In these competitive evaluation campaigns, the organizers provide annotated data sets for training in advance, which are used by participating teams from all over the world to fine-tune their systems. Later, test data sets are released for a limited period of time before the official results are provided by the participants. After a shared task has finished, these data sets are used to evaluate emerging new systems, thus enabling comparison between systems. The area of automatic fake news detection is relatively recent and, for this reason, this type of shared tasks has not become popular until very recently. A list of the most relevant shared tasks follows (as in the previous section, the target language of the tasks is English unless specified otherwise):

- The goal of the Fast & Furious Fact Check Challenge (<https://www.herox.com/fact-check/teams>, accessed on 3 June 2021), held in 2016, was to develop faster ways to check facts through automation. The challenge consisted in assigning a “truth rating” to each claim (there were 90 claims, 41 for training and 49 for testing). The small size of the set of claims prevents its use by today’s prevalent machine learning techniques.
- The Fake News Challenge FNC-1 (<http://www.fakenewschallenge.org/>, accessed on 3 June 2021) [113] held in 2017 aimed to determine the perspective (or stance) of a news article relative to a given headline. Therefore, despite its name, this task does not detect fake news but determines its stance. An article’s stance can either agree or disagree with the headline, discuss the same topic, or be completely unrelated. An existing data set for stance classification [91] was enhanced for this purpose, with 50 K labelled claim–article pairs, combining 300 claims with 2582 articles. The claims and the articles were curated and labeled by journalists.
- WSDM—Fake News Classification (<https://www.kaggle.com/c/fake-news-pair-classification-challenge/>, accessed on 3 June 2021) was a shared task organized within the framework of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM 2019). More than the detection of fake news, this task dealt with the detection of news that propagated or refuted other fake news, since given the title of a fake news article A and the title of a news article of interest B, participants were asked to classify B into one of three categories: “agreed” (B talks about the same fake news as A), “disagreed” (B refutes the fake news in A), and “unrelated” (B is unrelated to

- A). The training data set contained 320,767 news pairs in both Chinese and English, while the test data set contained 80,126 news pairs in both languages.
- Fake News Detection Challenge KDD 2020 (<https://www.kaggle.com/c/fakenewskdd2020>, accessed on 3 June 2021) was a shared task organized in the context of the Second International TrueFact Workshop: Making a Credible Web for Tomorrow in conjunction with SIGKDD 2020. The data set was composed of fake and true news.
 - Profiling Fake News Spreaders on Twitter (<https://pan.webis.de/clef20/pan20-web/author-profiling.html>, accessed on 3 June 2021) [114] was a shared task organized within CLEF 2020 in the context of PAN, a series of scientific events, and shared tasks on digital text forensics and stylometry. The particularity of this task is that it does not properly try to detect fake news but to detect whether a Twitter user is a potential propagator of fake news. The languages considered for this task were English and Spanish. The data set consisted of the last 100 tweets of 500 users (250 fake news spreaders and 250 true news spreaders) for each of the two languages.
 - UrduFake@FIRE2020 (<https://www.urdufake2020.cicling.org/>, accessed on 3 June 2021) [115] was a shared task organized within the framework of the 12th meeting of the Forum for Information Retrieval Evaluation (FIRE 2020). The data set used for the task was the Bend-The-Truth Urdu fake news data set, composed of 750 true news articles in the domains of technology, education, business, sports, politics, and entertainment, obtained from a variety of mainstream news websites predominantly in Pakistan, India, the United Kingdom, and the USA, and 550 fake news intentionally written by a group of professional journalists in the same domains and of the approximately same length as true news.
 - The CONSTRAINT 2021 shared task (<https://constraint-shared-task-2021.github.io/>, accessed on 3 June 2021) [116] was organized in the context of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT 2021) collocated with AAAI 2021. It encompassed a task for detecting fake news in English over a data set with real and fake news on COVID-19 [117] and a second task on hostile post detection in Hindi over a data set of 8200 hostile and non-hostile texts from various social media platforms such as Twitter, Facebook, or WhatsApp [118]. This latter task is a multi-label, multi-class classification problem where each post can belong to one or more of a set of classes: fake news, hate speech, offensive, defamation, and non-hostile posts.
 - FakeDeS (<https://sites.google.com/view/fakedes>, accessed on 3 June 2021) is the Fake News Detection in Spanish Shared Task established in 2021 under the umbrella of the Iberian Languages Evaluation Forum (IberLEF). The training data set for this task is The Spanish Fake News Corpus [110] described above. The testing corpus contains news related to COVID-19 and news from other Ibero-American countries.

4.4. Evaluation Measures

The most widely used measures to determine the performance of fake news detection systems are accuracy (Acc), precision (P), recall (R), F1, and Area Under the Curve (AUC). In the context of fake news, the articles annotated in the test data set are considered the ground truth. To compute the metrics, we must consider the number of true positives ($|TP|$), true negatives ($|TN|$), false positives ($|FP|$), and false negatives ($|FN|$) with respect to said ground truth, where

- A true positive is counted for each article predicted as fake news that is actually annotated as fake in the test set;
- A false positive is counted for each article predicted as fake news that is actually annotated as true news in the test set;
- A true negative is counted for each article predicted as true news that is actually annotated as true in the test set; and
- A false negative is counted for each article predicted as true news that is actually annotated as fake in the test set.

Accuracy refers to the percentage of articles that has been labeled correctly by the detection system, either as fake or true news, and is computed as indicated below in Equation (1). Precision refers to the percentage of fake news predicted that correspond to actual fake news and is computed as indicated in Equation (2). Recall refers to the percentage of total fake news items in the test set that are successfully recognized as fake by the detection system and is computed as indicated in Equation (3). F1 combines precision and recall by means of their harmonic mean, as shown in Equation (4). For all of these measures, the higher the value, the better the performance.

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |FP| + |TN| + |FN|} \quad (1)$$

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|} \quad (2)$$

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|} \quad (3)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The Receiver Operating Characteristics (ROC) curve provides a way of comparing the performance of several detection systems by looking at the trade-off in the False Positive Rate (FPR, also known as fall-out, equivalent to one minus the True Negative Rate or specificity) and the recall, which is also known as the True Positive Rate (TPR) or sensitivity in this context. FPR is computed as indicated in Equation (5). A ROC curve is drawn by plotting the FPR on the horizontal axis and the TPR along the vertical axis.

$$\text{FPR} = \text{fall-out} = \frac{|FP|}{|FP| + |TN|} \quad (5)$$

AUC is precisely the area under the ROC curve. An excellent system has an AUC close to 1 (it is able to perfectly distinguish between all fake and true news correctly), while a poor system has an AUC close to 0 (it would be considering all fake news as true and all true news as fake). AUC is more statistically consistent and more discriminating than accuracy [119], and it is usually applied in imbalanced classification problems, as is the case of fake news detection, where the number of ground truth fake news articles and true news articles have a very imbalanced distribution [12].

4.5. Summary

The 2016 US presidential election marked a turning point in the interest in fake news, both in society as a whole and in the scientific community in particular. This has led to the creation of a large number of resources, mainly in the form of data sets but also in the form of shared tasks. The negative part of all this is that the research community has not been able to establish standard elements (data sets and performance measures) that can be used by all authors involved in the field in order to allow an adequate comparison of the various approaches that have been developed over the last five years.

5. Sentiment Analysis as a Base Component for Text Analytics

In text analytics and NLP, subjective information usually refers to the fragments of natural language utterances that reveal opinions, feelings, points of view, and stances with respect to a given topic of interest. Automatically analyzing such utterances and understanding the sentiment that is expressed there is usually referred to as sentiment analysis, and it is interesting not only from a pure research point of view but also from an industry one. In this way, it is possible to process huge amounts of data to monitorize the views of society with respect to public matters, events, or products. SA started to become popular in the late 90s, when the first practical data sets for subjective classification were

introduced [120] and when the first paradigms that made it possible to create models to tackle the problem effectively were proposed.

On the one hand, Turney [121] introduced what are now commonly called lexicon-based methods. His approach consisted in automatically creating semantic orientation dictionaries that mapped words to subjectivity scores by computing the pointwise mutual information of adjectives and adverbs against the seed words “excellent” and “poor”. Word scores could be then added up to determine the overall sentiment of a given input. Many authors have followed this angle to create their SA systems. For instance, Esuli and Sebastiani [122] presented SentiWordNet, a lexical resource where every entry in WordNet [123] is assigned a positive, neutral, and negative score, which is generated by a set of ternary semi-supervised classifiers. Differently, Taboada et al. [124] and Brooke et al. [125] manually created semantic orientation dictionaries for common nouns, verbs, adverbs, adjectives, and intensifiers. They also presented a simple yet effective rule-based system (for English and Spanish) that dealt with relevant phenomena for polarity classification, such as negation or intensification. Thelwall et al. [126] presented SentiStrength, another lexical rule-based system for sentiment classification but focused on short texts while being easily adaptable to different languages. Hutto and Gilbert [127] enhanced a lexicon-based SA system with five generalizable rules that try to embody some writing conventions that humans use when expressing or emphasizing sentiment intensity. Vilares et al. [128] went one step further and proposed a rule-based system that was based on syntactic structures, and it is applicable to any language for which a dependency parser and a semantic orientation dictionary are available. Related to lexicon-based methods, other authors such as Cambria et al. [129,130] have focused instead on concept-level sentiment analysis, creating semantic graphs that connect concepts in order to capture their semantics.

On the other hand, subjective analysis can be also tackled using machine learning models, an angle that has been especially dominant since the general adoption of deep learning in NLP. Pang et al. [131] were among the first to use machine learning models for SA, and studied the use of n-grams of words and part-of-speech tags to train naive Bayes, maximum entropy, and support vector machine classifiers. Their setup served as the starting point for many authors to explore the usefulness of different types of features for SA, such as lexical [132], morphological [133], and syntactic features [134,135]. Kalchbrenner et al. [136], dos Santos et al. [137], and Severyn and Moschitti [138] were among the first to apply deep learning techniques over raw texts for SA, which has greatly reduced the need to manually engineer features thanks to models such as convolutional neural networks. In the same line, Socher et al. [139] introduced a novel model to compute the sentiment of texts with a tree-recursive neural network. To do so, they first created a sentiment treebank where constituents (phrases) were annotated by Amazon Mechanical turkers, which was then used to train a recursive deep model that learned to generate the sentiment of upper constituents by composing the sentiment of their children. In a related line, Tai et al. [140] showed how a tree-LSTM that exploited syntactic properties could obtain slight improvements over raw LSTM models for SA (as well as for other sequential tasks). Lately, the use of distant learning and semi-supervised learning has been used to obtain significant improvements in SA. For instance, Radford et al. [141] showed how pre-training an LSTM on a large data set of Amazon reviews at the character level (byte level) could obtain state-of-the-art results for SA on standard data sets by simply fine-tuning on a few supervised examples. Similarly, Devlin et al. [142] showed how BERT, the well-known bidirectional language model, obtained strong results for SA after the fine-tuning phase. For space reasons, we refer the reader to [143] for a complete review on deep learning methods for sentiment analysis.

SA models, methods, and techniques have been successfully applied in the context of text analytics; in applications such as the analysis of reviews on products and services [128,144]; in the analysis of social media posts in Twitter [145,146], Facebook [147], Instagram [148], etc.; in the detection of social spam to prevent normal users from being unfairly overwhelmed with unwanted or fake content via social media [149]; in the detection

or irony, sarcasm, and satire in formal and informal text [150,151]; in the detection of sexism, racism, bullying, harassment, and hate speech [152–156]; in influence and reputation analysis [157,158]; in political [159,160], social [161,162], and economic analysis [163,164]; in security monitoring [165]; and in health and well-being analyses [166,167].

6. Sentiment Analysis for Fake News Detection

As established in previous sections, the sentiment that is expressed and the strength with which it is expressed constitute a substantial element to reliably determine the degree of veracity of a news piece. Among the developments in the last decade, we can distinguish two types of approaches for detecting fake news as far as the use of SA is concerned. On the one hand, there is a set of approaches that take SA as the fundamental basis of their fake news detection strategy, which is usually complemented by the use of other information extracted from both news contents and the context of news spreading on social networks. These are the approaches to which we dedicate Section 6.1. On the other hand, we have a more numerous set of models in which the sentiment expressed in a news item is considered as a feature along with other features obtained from the text and the context of the news piece. Papers that have chosen this approach are the subject of Section 6.2.

To compile the papers discussed in this section, we took into account that, in general, computer science researchers prefer to publish their findings at scientific conferences over journals and that the coverage of these conferences is insufficient in Web of Science (<http://webofknowledge.com/>, accessed on 3 June 2021) and Scopus (<https://www.scopus.com/>, accessed on 3 June 2021) but quite good in Google Scholar (<https://scholar.google.es/>, accessed on 3 June 2021). Consequently, we searched for relevant articles on Google Scholar by using as search terms “fake news sentiment” and “fake news opinion”. We considered it relevant to note that, in this review, we prioritize presenting each related paper in detail over organizing the full research on the subject of the review. Thus, we follow the approach used in other reviews on fake news published in scientific journals in the last two years [4,6,19,23,25]. The other possibility would have been to follow the approach, less common in computer science than in other disciplines such as those related to healthcare, of carrying out a systematic review according to some guidelines, as it was the case of [18,20,26].

6.1. Fake News Detection Systems Based on SA

Diakopoulos et al. [168] presented an analytic tool designed to help journalists and media professionals to extract news value from large-scale aggregations of social media content around broadcast events and, in particular, to drive analysts to gather information from identified sources. For this purpose, they leveraged four types of automatic content analysis: relevance, uniqueness, sentiment, and keyword extraction. For sentiment analysis, they followed a two-step approach. As a first step, they ran a simple classifier based on a lexicon of words that classified texts based on whether they were carrying subjective information. In a second step, they applied a supervised machine learning classifier based on n-grams of length up to four. The combined classifier resulted in a five-fold cross-validated accuracy of 62.4%, which was sufficient for giving an overall impression of the sentiment but still failed on difficult cases involving sarcasm or slang. Their tool provided a visual representation of aggregated sentiment in the form of a sentiment timeline. They evaluated the tool on 101,285 tweets about the State of the Union address by US President Barack Obama in early 2010 and they found that sentiment analysis in conjunction with the magnitude of the social media response to different quotes, topics, or issues were the most useful analytic indicators for journalistic inquiry.

AlRubaiian et al. [169] also used SA to identify implausible content in Twitter written in Arabic in order to prevent the proliferation of fake or malicious information, since they considered that sentiment gives a measurement about the user behavior, which led to high precision in credibility analyses. In their system, sentiment weighted 30% with respect to all of the features that were considered. SA, along with topic classification, was also the

basis of the system by Chatterjee and Agarwal [170] to determine the credibility of tweets written in English and Hinglish (a hybrid use of English and Hindi).

Dey et al. [171] applied several NLP methods (part-of-speech tagging, named-entity recognition, sentiment analysis) to a set of 200 tweets referred to the 2016 US Presidential Election. They found that credible tweets had mostly positive or neutral polarity, while tweets with fake content had a strong tendency towards negative sentiment. However, their data set was too small to obtain conclusive results. Bhutani et al. [172] based their fake news detector on sentiment analysis under the hypothesis that the sentiment that was put forth on writing a news article would serve as a pivotal deciding factor in the process of characterizing the news into fake or real. They applied a Naive-Bayes classifier to determine the sentiment of texts and then used it as a main feature of Multinomial Naive-Bayes and Random Forest classifiers for detecting fake news, with the latter obtaining the best results.

Ajao et al. [173] also exploited the hypothesis that there is a relation between fake news and the sentiment of the text posted online. This hypothesis was statistically tested on a corpus of rumors [174] and was confirmed later in experimental results to compare several classic and deep learning classifiers using sentiment against a deep learning baseline considering only text features. The best results for fake news detection were obtained by sentiment-aware classifiers, an SVM in the case of classic models and an LSTM with hierarchical attention in the case of the deep learning ones.

Cui et al. [175] found statistical evidence on the FakeNewsNet data set that the sentiment polarity of comments under fake news was greater than under real news. As a consequence, they decided to incorporate users' latent sentiments into an end-to-end deep embedding framework for detecting fake news. They used three neural networks to deal with news images, news text, and user profiles, whereas an adversarial mechanism was introduced to preserve semantic similarity, and enforce representation consistency between the text and image. Finally, they modeled users' sentiment to incorporate it into the proposed framework. A novelty of this work was the use of adversarial learning to find semantic correlations between different modalities in news contents. The resulting fake news detection system outperformed other classical and deep learning-based classifiers. Ablation experiments showed that the component that contributed the most to the performance of the system was sentiment analysis.

Del Vicario et al. [176] introduced a framework for early warning on possible misinformation targets on social media. They compiled a data set with real news from Facebook posts from Italian official newspapers and with fake news from Facebook posts taken from Italian sites known for disseminating misinformation. For each post, they extracted the entities associated with the textual content and the polarity of the sentiment expressed in the text. For each entity, they calculated the "presentation distance" (the absolute difference between the maximum and the minimum values among the sentiment scores of all posts containing the entity) and the "mean response distance" (the absolute difference between the mean sentiment score on the posts containing the entity and the mean sentiment score on their comments). From these two values, they set the controversy and the perception of the entity according to empirically derived thresholds. They noticed that presentation distance was a good indicator of the attention received by an entity in terms of likes and comments and that both controversial and captivating entities were much more present in fake news, thus highlighting the potential of such properties in identifying topics that were likely to be subject to misinformation. Finally, they used these measures to derive sentiment-based features which, together with others based on text properties (e.g., number of characters, words, etc.) and user behavior (e.g., number of comments, likes, etc.), fed several classic machine-learning classifiers to detect fake news. The best performance was attained by a logistic regression classifier.

Fake health news pieces have some characteristics that make them harder to detect than fake news on other domains. For example, they may mislead the reader by stating association as causation or by mixing up absolute risk and relative risk, which only require

minor modifications based on the true information. Dai et al. [56] conducted an exploratory analysis to understand the characteristics of the data sets for health fake news detection, to analyze useful patterns, and to validate the quality of the FakeHealth data sets. With respect to social engagement on health news, they found that replies towards real news were more positive. In the same domain, Anoop et al. [55] targeted the detection of fake health news within online media sources that resembled traditional newspapers, given that the information was in the form of articles with some trustworthy information and abundant emotion-oriented narrative. For this reason, they based their approach for fake news detection on the different kinds of affective characteristics that were displayed on fake and true health news articles. Emotion features were extracted from a lexicon to feed classical and deep learning classifiers, and the results showed that emotion information increased performance for all classifiers. They also conducted preliminary experiments on detecting fake news on COVID-19, where they found a significant presence of emotional content in the narratives, indicating the applicability of emotion-oriented detection to identify fake news about this pandemic.

Zhang et al. [112] considered that most existing work on fake news was based on the emotional signals of content conveyed by the publishers but rarely focused on the emotions of comments aroused in the crowd, even when viral spread is fueled by the evocation of high-arousal emotions. Precisely, they explored whether emotions in news comments and their relationship to those in the content itself were helpful for fake news detection. They tested the approach using various deep learning-based classifiers on data sets in English and Chinese. The results on a Chinese fake news data set were good, while the results on the English data set were quite low, probably because the data set was originally designed for the detection of rumors, not fake news.

6.2. SA as a Feature for Fake News Detection Systems

Before fake news became a first-order issue, there was already some work in the literature that focused on determining the credibility of information circulating on social networks. In this line, Castillo et al. [84] focused on automatic methods for assessing the credibility of sets of tweets that spread information about a news event. They considered information from official and reputable sources as valuable information that other users synthesize and elaborate to produce derived interpretations in a continuous process. For their experiments, they collected 747 sets of tweets spreading news, each of one was manually tagged as “almost certainly true”, “likely to be false”, “almost certainly false”, and “I can’t decide”. They noted that features based on sentiment analysis were very relevant for assessing credibility, as 3 out of the 10 best-performing features were sentiment-related ones: average sentiment score, number of positive sentiment words, and number of negative sentiment words. They also observed that tweets exhibiting positive sentiment were more related to non-credible information while those with negative sentiment tended to be more related to credible information.

Ross and Thirunarayan [177] differed from [84] in that, instead of trying to determine the credibility of a tweet, they focused on ranking a collection of tweets by credibility and newsworthiness. In their base set of features, there were sentiment features, “has a happy emoticon” and “has a sad emoticon”, and a sentiment score. In addition, they considered two features that aimed to capture when the sentiment of a tweet matched the overall sentiment of the topic hypothesizing that tweets that had similar sentiments to the topic would be credible, while tweets with dissimilar sentiments could mean that the tweet was non-credible or not newsworthy. They obtained features not only from the text of the tweet but also from the description of the authors of the tweet: the number of positive, negative, and curse words in the user description.

Popat et al. [102] addressed the problem of assessing the credibility of arbitrary textual claims that were expressed freely in an open-domain setting by automatically finding sources in news and social media. These sources were then fed into a logistic regression classifier in order to determine whether the claim was true or fake. A key element of their

approach was the analysis of the style in which a claim was reported in an article, as it was assumed that a true claim was to be reported in an objective and unbiased language [178]. They captured the linguistic style by means of a set of lexicon-based features such as a list of positive and negative opinionated words; lists of assertive, factive, and report verbs; hedge words; implicative words; and discourse markers. This approach assumed that ample evidence or counter-evidence could be easily retrieved from a static snapshot of the web, but this is not true for newly emerging claims with sparse presence on the web. To overcome this limitation, Papat et al. [179] proposed to enhance the approach by determining the stance, reliability, and trend of retrieved sources of evidence or counter-evidence, which were used with the features already defined in [102] to feed a CRF classifier.

Hassan et al. [180] described a semi-automatic detection system that monitored live discourses, social media, and news to catch factual claims that were translated automatically into queries against a curated repository of fact-checks. For some claims, humans were brought into the loop, and in that case, the automated system assisted them in understanding and verifying the claims. The system was trained on a set of US general election presidential debates where each sentence was annotated as “Non-Factual”, “Unimportant Factual”, or “Check-worthy Factual”. For each sentence, they extracted as features its sentiment polarity, its length, its bag of words, the number of occurrences of every part of speech, and its named entities [181]. All of this resulted in a set of 6615 features, among which sentiment turned out to be the third most relevant one.

Varol et al. [182] studied the development of computational methods for the early detection of information campaigns that can be used to spread fake news, propaganda, or financial market manipulation. In particular, they tried to determine whether a Twitter hashtag was promoted based on information that would be available even in cases where the nature of a trend is unknown. This is a difficult task because a minority of promoted conversations is blended into a majority of organic content. Additionally, promoted hashtags may preexist the moment in which they are given the promoted status and may have originated in an entirely regular form, thus displaying features that are largely indistinguishable from those of other regular hashtags about the same topic until the moment of promotion. They proposed to use 487 features extracted from network structure and dissemination patterns, language, content and sentiment information, timing signals, and user meta-data. Sentiment-based features included sentiment polarity and strength, positive and negative emoticons, happiness score, arousal, valence and dominance scores, and emotion score. From these, only measures related to emoticons were among the top 10 features for experiments performed with machine learning classifiers.

Content differences between fake and real political news were studied by Horne and Adali [10]. They kept track of the sentiment polarity for each sentence, stylistic features (the number of times each part of speech appears in an article, the number of stopwords, punctuation, quotes, negations, informal/swear words, interrogatives, and words that appear in all capital letters), word complexity (number of syllables in words, ratio of unique words, and number of common and specialized words), and sentence complexity (number of words per sentence, depth of the sentence’s syntax tree, and depths of the syntax trees for noun and verb phrases). They found that positive and negative sentiments were statistically significant features to differentiate the body text of real and fake news in the Silverman’s BuzzFeed Political News Data set. This did not happen when only news headlines were considered, probably due to their short length.

Rashkin et al. [99] studied the language of news media in the context of political fact-checking and fake news detection. For this purpose, they estimated the use of strongly and weakly subjective words with a sentiment lexicon under the hypothesis that subjective words were used to dramatize or sensationalize news stories. Other types of lexical cues they considered were hedge words, intensifiers, comparatives, superlatives, action adverbs, manner adverbs, and modal adverbs. They found that words that can be used to exaggerate (subjective words, superlatives, and modal adverbs) were all used more by fake news. However, the additional features were only useful with classical classifiers, since the

improvements for classifiers based on deep neural networks were negligible with respect to using the text as the only input feature.

Vosoughi et al. [63] analyzed true and fake news stories disseminated on Twitter from 2006 to 2017. They found that fake stories inspired the emotions of fear, disgust, and surprise in replies, whereas true stories inspired anticipation, sadness, joy, and trust. They conclude that the emotions expressed in reply to fake news may illuminate additional factors that inspire people to share false news. Although emotion analysis is not the same as SA, they are closely related because both analyze the subjective content expressed in a text, and the classification models used are very similar in both cases, with the biggest difference being in the set of classes with which the texts to be processed are annotated. For this reason, we have included this article in the analysis.

Yang et al. [183] analyzed the BS Detector data set and concluded that the polarity of sentiment in true and fake news was different, with fake news tending towards negative sentiment. The standard deviation of fake news on negative sentiment was also larger than that of real news, thus some of the fake news would have very strong negative sentiment. With respect to the use of language in true and fake news, they observed that fake news had fewer words and sentences and that the variance of these values in true news was much smaller than that in fake news; true news had fewer question marks; fake news had a much larger ratio of capital letters; the median of negations in fake news was much smaller; fake news had fewer first-person and second-person and more third-person pronouns; true news had more lexical diversity; and there were differences in the usage of words in titles of fake and true news. Besides these explicit features, the approach of [183] was based on the consideration that there exist hidden patterns in the words and images used in fake news that can be captured with a set of latent features extracted via the multiple convolutional layers in a deep neural network. The novelty of this approach was that it proposed a unified model to analyze both text and images from fake news. In particular, they used two parallel CNNs to extract latent features from both textual and visual information. Then, explicit and latent features were projected into the same feature space.

Reis et al. [184] extracted a large set of features from news content by using language processing techniques as well as from news sources (bias, credibility and trustworthiness, and domain location) and from the environment (number of likes, shares and comments, and the rate at which comments are posted). One of the context features was the subjectivity and sentiment scores of the news text. All of these features were fed to several classic classifiers, with Random Forest and XGBoost obtaining the best performance on the BuzzFace data set. From their results, the authors observed that it was possible to choose a threshold so as to correctly classify almost all of fake news (TPR close to 1) while misclassifying 40% of the true news (FPR of 0.4), and they considered that this could be useful in assisting fact-checkers in identifying stories worth investigating.

Shu et al. [111] analyzed the FakeNewsNet data set, finding that people express their emotions or opinions toward fake news through social media posts such as skeptical opinions and sensational reactions, with real news having a larger proportion of neutral replies over positive and negative replies, whereas fake articles have a bigger ratio of negative sentiment. In their preliminary experiments to classify fake news, they used base features from the text only, so they did not provide information about the impact that SA has on the detection of fake news in this data set.

7. Discussion

We have seen how SA can be used to improve the effectiveness of fake news detection systems in a number of ways. Tables 1–3 show a summary of the most notable characteristics of the systems discussed in the previous section. In these tables, CRF stands for Conditional Random Field, SVM stands for Support Vector Machine, MaxEnt stands for Maximum Entropy, LSTM stands for Long Short-Term Memory, KNN stands for k-Nearest Neighbors, TI-CNN stands for Text and Image information-based Convolutional Neural Network [183], LSTM-HAN stands for Long Short-Term Memory with Hierarchical Atten-

tion Networks [185], BiGRU stands for Bidirectional Gated Recurrent Unit, BERT stands for Bidirectional Encoder Representations from Transformers [142], NileTMRG stands for Nile Text-Mining Research Group [186], and HSA-BLSTM stands for Hierarchical Social Attention Bidirectional Long Short-Term Memory [187].

Some of the studies did not show experimental performance results on the fake news detection task for various reasons. Such systems are listed in Table 1: Diakopoulos et al. (2010) [168] performed a qualitative assessment on perceived utility rather than a quantitative analysis; Chatterjee and Agarwal (2016) [170] presented a qualitative assessment of results on a limited set of tweets; the purpose of Ross and Thirunarayan's [177] system was to rank tweets based on their credibility, which means that the performance measures used were not comparable to those commonly used in fake news detection systems; Hassan et al. [180] did not provide comprehensive results on the detection of false claims but provided a comparison of the detected percentages against journalistic news verification media such as CNN and PolitiFact; and Vosoughi et al. [63] did not present measures of news classification between fake and true news but instead presented statistical analyses with respect to the assumptions they made regarding the speed of news spread on Twitter.

Table 1. Main characteristics of fake news detection systems using SA: systems that do not provide quantitative performance results on the task.

| Reference | Language | SA Method | Detection Method | Data Set |
|-------------------------------------|----------------------|----------------------------------|-------------------|-------------------------------------|
| Diakopoulos et al. (2010) [168] | English | Lexicon + ngram-based classifier | N/A ¹ | Ad-hoc from Twitter |
| Chatterjee and Agarwal (2016) [170] | English and Hinglish | SVM | Pipeline LDA + SA | Ad-hoc from Twitter |
| Ross and Thirunarayan (2016) [177] | English | Ranking SVM | N/A | Ad-hoc from Twitter |
| Hassan et al. (2017) [180] | English | Commercial tool (Alchemy) | SVM | Ad-hoc from US Presidential debates |
| Vosoughi et al. (2018) [63] | English | Lexicon-based | N/A | Ad-hoc from Twitter |

¹ N/A = Not applicable or not available (information not provided in the paper).

Tables 2 and 3 show the systems for which performance measures were provided. We can see how the first models used especially created data sets to carry out the experiments described in each of the papers. Later, as time progressed, more and more data sets were used that could be considered standard in the sense that they are available for use by other researchers to corroborate the results or to test their own approaches to the task. We can also observe how, until 2019, most of the works tested a single machine learning system for the final phase of the detection system, while from that year on, it is common for a given paper to report results for different classifiers. To put these results in context with those obtained with other fake news detection methods not involving SA, Oshikawa et al. [24] reported that the best performing system on the LIAR data set achieved an accuracy of 0.457. In the case of FEVER, the best system achieved an accuracy of 0.647. In contrast, the accuracy of the best-performing systems increases to 0.944 for Buzzfeed Political News Data and 0.938 for the PolitiFact portion of FakeNewsNet. With regard to FakeNewsNet, Zhou and Zafarini [23] provided an analysis of the evolution of performance as lexical, syntactic, semantic, and discourse analysis features were considered.

It is difficult to compare results across systems due to the use of an abundant variety of different data sets and various performance measures. While many of the systems report the F1 score value along with that of precision P and recall R, this is because F1 is actually an aggregate measure of P and R. On the other hand, the systems that provide the value of the accuracy Acc do not report the value of F1 and vice versa. Only a couple of systems report the value of Acc and AUC, one provides AUC and F1 and another one shows Acc

and F1. In practice, this diversity of measurements is not so relevant because the only systems that can be compared in the same evaluation data set are those created by the same authors. From a qualitative point of view, it is not uncommon that, although the results of many articles may coincide in many of their conclusions, they present discrepancies in certain specific aspects. We consider this to be due to the relative youth of the fake news detection field. With the organization of dedicated shared tasks in 2020 and 2021 and the creation of large data sets in recent years, we expect that, in the near future, it will be increasingly common to see systems that are evaluated on these new standard corpora.

For the SA methods used in these systems, a measure of the performance of the SA systems that have been applied is not normally given. This prevents us from knowing if an improvement in SA performance induces a significant improvement in final fake news detection performance. Furthermore, most papers use relatively simple, lexicon-based SA systems. The models that constitute the state-of-the-art in SA are based on machine learning. These systems are more expensive to build and require more computational resources than lexicon-based systems. It would be desirable to know if this cost increase is worth it, that is, if it would be reflected in an improvement in the effectiveness of a fine fake news detection system.

Table 2. Main characteristics of fake news detection systems using SA: systems that provide quantitative performance results on the task (part 1 of 2).

| Reference | Language | SA Method | Detection Method | Data Set | Performance |
|-------------------------------|----------|---------------|----------------------------|-------------------------------------|--|
| Castillo et al. (2011) [84] | English | Lexicon-based | J48 decision tree | Ad-hoc from Twitter | P = 0.861; R = 0.860; F1 = 0.860 |
| AlRubaian et al. (2015) [169] | Arabic | Lexicon-based | Naive-Bayes | Ad-hoc from Twitter | P = 0.8624; R = 0.988; F1 = 0.926; Acc = 0.903 |
| Popat et al. (2016) [102] | English | Lexicon-based | Logistic regression | Snopes | Acc = 71.96; AUC = 0.80 |
| Popat et al. (2017) [179] | English | Lexicon-based | CRF | Snopes | Acc = 84.02; AUC = 0.86 |
| Horne and Adali (2017) [10] | English | Lexicon-based | SVM | Silverman's Buzzfeed Political News | Acc = 0.77 |
| | | | | Ad-hoc (news articles) | Acc = 0.71 |
| Rashkin et al. (2017) [99] | English | Lexicon-based | MaxEnt | Fact Checking | F1 = 0.55 |
| | | | LSTM | | F1 = 0.56 |
| Varol et al. (2017) [182] | English | Lexicon-based | KNN | Ad-hoc from Twitter | Acc = 0.97; F1 = 0.81 |
| Dey et al. (2018) [171] | English | Lexicon-based | KNN | Ad-hoc from Twitter | Acc = 0.66 |
| Yang et al. (2018) [183] | English | N/A | TI-CNN | BS Detector | P = 0.9220; R = 0.9277; F1 = 0.9210 |
| Bhutani et al. (2019) [172] | English | Naive-Bayes | Random Forest | LIAR | AUC = 0.63 |
| | | | | Fake vs. real news project | AUC = 0.88 |
| Ajao et al. (2019) [173] | English | Lexicon-based | SVM | Rumors [174] | Acc = 0.86; P = 0.86; R = 0.86; F1 = 0.86 |
| | | | LSTM-HAN | | Acc = 0.86; P = 0.86, R = 0.82; F1 = 0.84 |
| Cui et al. (2019) [175] | English | Rule-based | Ad-hoc deep neural network | FakeNewsNet PolitiFact | F1 = 0.7724 |
| | | | | FakeNewsNet GossipCop | F1 = 0.8042 |

Table 3. Main characteristics of fake news detection systems using SA: systems that provide quantitative performance results on the task (part 2 of 2).

| Reference | Language | SA Method | Detection Method | Data Set | Performance |
|---------------------------------|----------|---------------------------------------|---------------------|----------------------|---|
| Del Vicario et al. (2019) [176] | Italian | Commercial tool (Dandelion API) | Linear Regression | Ad-hoc from Facebook | P = 0.90; R = 0.90; FPR = 0.11; F1 = 0.90 |
| | | | Logistic Regression | | P = 0.91; R = 0.91; FPR = 0.08; F1 = 0.91 |
| | | | KNN | | P = 0.87; R = 0.87; FPR = 0.13; F1 = 0.87 |
| | | | Decision tree | | P = 0.89; R = 0.89; FPR = 0.12; F1 = 0.89 |
| Reis et al. (2019) [184] | English | Rule-based | KNN | BuzzFace | AUC = 0.80; F1 = 0.75 |
| | | | Naive Bayes | | AUC = 0.72; F1 = 0.75 |
| | | | Random Forest | | AUC = 0.85 ; F1 = 0.81 |
| | | | SVM | | AUC = 0.79 ; F1 = 0.76 |
| | | | XGBoost | | AUC = 0.86 ; F1 = 0.81 |
| Anoop et al. (2020) [55] | English | Lexicon-based | Naive Bayes | HWB | Acc = 0.790 |
| | | | KNN | | Acc = 0.925 |
| | | | SVM | | Acc = 0.900 |
| | | | Random Forests | | Acc = 0.840 |
| | | | Decision Tree | | Acc = 0.940 |
| | | | AdaBoost | | Acc = 0.965 |
| | | | CNN | | Acc = 0.910 |
| | | | LSTM | | Acc = 0.920 |
| Zhang et al. (2021) [112] | English | Lexicon-based + Commercial (NVIDIA) | BiGRU | RumourEval-19 [188] | F1 = 0.340 |
| | | | BERT | | F1 = 0.346 |
| | | | NileTMRG | | F1 = 0.342 |
| | Chinese | Lexicon-based + Commercial (Baidu AI) | BiGRU | Weibo-20 | F1 = 0.855 |
| | | | BERT | | F1 = 0.867 |
| | | | HSA-BLSTM | | F1 = 0.908 |

There are a number of challenges that need to be addressed in the near future in the fields of SA and fake news detection:

- **Multilingualism.** Most of the work published on detecting fake news has been conducted on documents written in English. However, as this is a global issue, it is imperative to have systems that work in as many languages as possible. It was only recently that approaches that are able to deal with fake news for several languages have appeared [189,190]. For this purpose, effective multilingual SA methods may be applied [128,145].
- **Multimedia content,** particularly image and video, is becoming increasingly important on social media. Fake news are also increasingly accompanied by this type of content, so it will be necessary to enrich detection systems with multimedia SA methods [191].
- **The most difficult fake news to detect** are those in which falsehood has been subtly introduced, for example, expanding an authentic news piece with the addition of fake data or slightly modifying an authentic news story [192]. In this case, aspect-based SA [193] and adversarial training [194] can be of great help.
- **High-performance AI systems,** particularly those based on deep learning, behave similar to black boxes that provide good results but can hardly justify a given output in a human-understandable way. The creation of explainable AI systems [195] is becoming more and more important, and therefore, it is necessary to add mechanisms both to the SA methods used and to the resulting fake news detection systems.

- It is known that NLP algorithms and resources may have inadvertently introduced biases [196,197]. This also applies to SA systems [198]. Adding algorithmic biases to psychological and sociological biases already present in people [199,200] could call into question the usefulness of automated fake news detection systems in the future. Therefore, we must add bias mitigation mechanisms in SA systems in order to avoid giving more or less credibility to a news item depending on the gender, race, geographic origin, religion, or any other personal circumstance of the writer or the people mentioned in the text.

8. Conclusions

The recent rise in the spread and social influence of fake news, driven by the popularization of social networks, has motivated a surge of interest in their automated detection. Since fake news tend to be written with the intent of conveying strong sentiments towards a given subject, sentiment analysis has proven to be a useful tool in the fake news detection toolbox, both when applied to news items themselves and to related information such as user comments.

In this article, we reviewed the field of fake news detection from the specific point of view of how sentiment analysis is being used to tackle the problem. We have seen that it has been proven useful in a diverse range of systems, both as a core component or as a source of auxiliary features. Direct comparison between systems and approaches is so far difficult due to the wide range of data sets used, many of them ad hoc, but this problem is on track to being solved with the recent appearance of publicly available data sets and shared tasks.

Thus, we can say that the research field of fake news detection (and in turn, the application of sentiment analysis for this purpose) is currently in its transition from infancy to maturity. In this stage, the most pressing challenges in our view involve the need to guarantee the fairness, accountability, and transparency of systems (ensuring that results are explainable and free from harmful biases); the support for multilingualism and multimedia content; and the detection of fake news generated by subtly modifying authentic stories or by using text-generation algorithms.

Author Contributions: Conceptualization, M.A.A., D.V., C.G.-R., and J.V.; methodology, M.A.A., D.V., C.G.-R., and J.V.; formal analysis, M.A.A., D.V., C.G.-R., and J.V.; investigation, M.A.A., D.V., C.G.-R., and J.V.; resources, M.A.A., D.V., C.G.-R., and J.V.; writing—original draft preparation, M.A.A., D.V., C.G.-R., and J.V.; writing—review and editing, M.A.A., D.V., C.G.-R., and J.V.; supervision, M.A.A., D.V., C.G.-R., and J.V.; project administration, M.A.A., D.V., C.G.-R., and J.V.; funding acquisition, M.A.A., D.V., C.G.-R., and J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by FEDER/Ministerio de Ciencia, Innovación y Universidades — Agencia Estatal de Investigación through the ANSWERASAP project (TIN2017-85160-C2-1-R); and by Xunta de Galicia through a Competitive Reference Group grant (ED431C 2020/11). CITIC, as Research Center of the Galician University System, is funded by the Consellería de Educación, Universidade e Formación Profesional of the Xunta de Galicia through the European Regional Development Fund (ERDF/FEDER) with 80%, the Galicia ERDF 2014-20 Operational Programme, and the remaining 20% from the Secretaría Xeral de Universidades (ref. ED431G 2019/01). David Vilares is also supported by a 2020 Leonardo Grant for Researchers and Cultural Creators from the BBVA Foundation. Carlos Gómez-Rodríguez has also received funding from the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation programme (FASTPARSE, grant No. 714150).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The BBVA Foundation accepts no responsibility for the opinions, statements, and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors.

References

1. Shearer, E.; Mitchell, A. News Use Across Social Media Platforms in 2020. 2021. Available online: <https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/> (accessed on 3 June 2021).
2. Herson, P. Disinformation and misinformation through the internet: Findings of an exploratory study. *Gov. Inf. Q.* **1995**, *12*, 133–139. [CrossRef]
3. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and Resolution of Rumours in Social Media: A Survey. *ACM Comput. Surv.* **2018**, *51*, 32:1–32:36. [CrossRef]
4. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
5. Zannettou, S.; Sirivianos, M.; Blackburn, J.; Kourtellis, N. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *ACM J. Data Inf. Qual.* **2019**, *11*, 10:1–10:37. [CrossRef]
6. Sharma, K.; Qian, F.; Jiang, H.; Ruchansky, N.; Zhang, M.; Liu, Y. Combating Fake News: A Survey on Identification and Mitigation Techniques. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 21:1–21:42. [CrossRef]
7. Guerini, M.; Staiano, J. Deep Feelings: A Massive Cross-Lingual Study on the Relation between Emotions and Virality. In Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, 18–22 May 2015; Companion Volume; Gangemi, A., Leonardi, S., Panconesi, A., Eds.; ACM: New York, NY, USA, 2015; pp. 299–305. [CrossRef]
8. Dickerson, J.P.; Kagan, V.; Subrahmanian, V.S. Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014, Beijing, China, 17–20 August 2014; Wu, X., Ester, M., Xu, G., Eds.; IEEE Computer Society: Washington, DC, USA, 2014; pp. 620–627. [CrossRef]
9. Chen, Y.; Conroy, N.J.; Rubin, V.L. Misleading Online Content: Recognizing Clickbait as “False News”. In Proceedings of the 2015 ACM Workshop on Multimodal Deception Detection, WMDD@ICMI 2015, Seattle, WA, USA, 13 November 2015; Abouelenien, M., Burzo, M., Mihalcea, R., Pérez-Rosas, V., Eds.; ACM: New York, NY, USA, 2015; pp. 15–19. [CrossRef]
10. Horne, B.D.; Adali, S. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. In Proceedings of the Workshops of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), Montreal, QC, Canada, 15–18 May 2017; An, J., Kwak, H., Benevenuto, F., Eds.; AAAI Press: Palo Alto, CA, USA, 2017; Volume AAAI Technical Report WS-17-17: News and Public Opinion, pp. 759–766.
11. Conroy, N.J.; Rubin, V.L.; Chen, Y. Automatic deception detection: Methods for finding fake news. In *Information Science with Impact: Research in and for the Community—Proceedings of the 78th ASIS&T Annual Meeting, ASIST 2015, St. Louis, MO, USA, 6–10 October 2015*; Wiley: Hoboken, NJ, USA, 2015; Volume 52, pp. 1–4. [CrossRef]
12. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor.* **2017**, *19*, 22–36. [CrossRef]
13. Shu, K.; Wang, S.; Lee, D.; Liu, H. Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements. In *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*; Shu, K., Wang, S., Lee, D., Liu, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–19. [CrossRef]
14. Shu, K.; Liu, H. Detecting Fake News on Social Media. In *Synthesis Lectures on Data Mining and Knowledge Discovery*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2019; Volume 18.
15. Hussein, D.M.E.D.M. A survey on sentiment analysis challenges. *J. King Saud Univ. Eng. Sci.* **2018**, *30*, 330–338. [CrossRef]
16. Flekova, L.; Preotiuc-Pietro, D.; Ruppert, E. Analysing domain suitability of a sentiment lexicon by identifying distributionally bipolar words. In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2015, Lisbon, Portugal, 17 September 2015; Balahur, A., der Goot, E.V., Vossen, P., Montoyo, A., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2015; pp. 77–84. [CrossRef]
17. Thorne, J.; Vlachos, A. Automated Fact Checking: Task Formulations, Methods and Future Directions. In Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, NM, USA, 20–26 August 2018; Bender, E.M., Derczynski, L., Isabelle, P., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 3346–3359.
18. Elhadad, M.K.; Li, K.F.; Gebali, F. Fake News Detection on Social Media: A Systematic Survey. In Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, PACRIM 2019, Victoria, BC, Canada, 21–23 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8. [CrossRef]
19. Bondielli, A.; Marcelloni, F. A survey on fake news and rumour detection techniques. *Inf. Sci.* **2019**, *497*, 38–55. [CrossRef]
20. da Silva, F.C.D.; Vieira, R.; Garcia, A.C. Can Machines Learn to Detect Fake News? A Survey Focused on Social Media. In Proceedings of the 52nd Hawaii International Conference on System Sciences, HICSS 2019, Grand Wailea, Maui, HI, USA, 8–11 January 2019; Bui, T., Ed.; ScholarSpace: Honolulu, HI, USA, 2019; pp. 1–8.
21. Klyuev, V. Fake News Filtering: Semantic Approaches. In Proceedings of the 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 29–31 August 2018; pp. 9–15. [CrossRef]
22. Collins, B.; Hoang, D.T.; Nguyen, N.T.; Hwang, D. Fake News Types and Detection Models on Social Media A State-of-the-Art Survey. In Proceedings of the Intelligent Information and Database Systems-12th Asian Conference, ACIIDS 2020, Phuket, Thailand, 23–26 March 2020; Companion Proceedings; Sitek, P., Pietranik, M., Krótkiewicz, M., Srinilta, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1178, pp. 562–573. [CrossRef]

23. Zhou, X.; Zafarani, R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Comput. Surv.* **2020**, *53*, 109:1–109:40. [CrossRef]
24. Oshikawa, R.; Qian, J.; Wang, W.Y. A Survey on Natural Language Processing for Fake News Detection. In Proceedings of the 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, 11–16 May 2020; Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., et al., Eds.; European Language Resources Association: Paris, France, 2020; pp. 6086–6093.
25. Zhang, X.; Ghorbani, A.A. An overview of online fake news: Characterization, detection, and discussion. *Inf. Process. Manag.* **2020**, *57*, 102025. [CrossRef]
26. de Souza, J.V.; Gomes, J., Jr.; de Souza Filho, F.M.; de Oliveira Julio, A.M.; de Souza, J.F. A systematic mapping on automatic classification of fake news in social media. *Soc. Netw. Anal. Min.* **2020**, *10*, 48. [CrossRef]
27. Antonakaki, D.; Fragopoulou, P.; Ioannidis, S. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Syst. Appl.* **2021**, *164*, 114006. [CrossRef]
28. Allcott, H.; Gentzkow, M. Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **2017**, *31*, 211–36. [CrossRef]
29. Wardle, C. Fake News. It's Complicated. 2017. Available online: <https://firstdraftnews.org/articles/fake-news-complicated/> (accessed on 3 June 2021).
30. Tandoc, E.C., Jr.; Lim, Z.W.; Ling, R. Defining “Fake News”. *Digit. J.* **2018**, *6*, 137–153. [CrossRef]
31. Golbeck, J.; Mauriello, M.L.; Auxier, B.; Bhanushali, K.H.; Bonk, C.; Bouzaghrane, M.A.; Buntain, C.; Chanduka, R.; Chekalos, P.; Everett, J.B.; et al. Fake News vs Satire: A Dataset and Analysis. In Proceedings of the 10th ACM Conference on Web Science, WebSci 2018, Amsterdam, The Netherlands, 27–30 May 2018; Akkermans, H., Fontaine, K., Vermeulen, I., Houben, G., Weber, M.S., Eds.; ACM: New York, NY, USA, 2018; pp. 17–21. [CrossRef]
32. Scheufele, D.A.; Krause, N.M. Science audiences, misinformation, and fake news. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 7662–7669. [CrossRef]
33. Field-Fote, E.E. Fake News in Science. *J. Neurol. Phys. Ther.* **2019**, *43*, 139–140. [CrossRef]
34. Taddicken, M.; Wolff, L. ‘Fake News’ in Science Communication: Emotions and Strategies of Coping with Dissonance Online. *Media Commun.* **2020**, *8*, 206–217. [CrossRef]
35. Kedar, H.E. Fake News in Media Art: Fake News as a Media Art Practice vs. Fake News in Politics. *Postdigit. Sci. Educ.* **2020**, *2*, 132–146. [CrossRef]
36. Ruzicka, V.; Kang, E.; Gordon, D.; Patel, A.; Fashimpaur, J.; Zaheer, M. The Myths of Our Time: Fake News. *arXiv* **2019**, arXiv:1908.01760.
37. Rapoza, K. Can ‘Fake News’ Impact The Stock Market? *Forbes* **2017**. Available online: <https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/> (accessed on 3 June 2021).
38. Clarke, J.; Chen, H.; Du, D.; Hu, Y.J. Fake News, Investor Attention, and Market Reaction. *Inf. Syst. Res.* **2020**, Forthcoming. [CrossRef]
39. Kogan, S.; Moskowitz, T.J.; Niessner, M. *Fake News in Financial Markets*; Social Science Research Network (SSRN): Rochester, NY, USA, 2020. [CrossRef]
40. Domenico, G.D.; Sit, J.; Ishizaka, A.; Nunan, D. Fake news, social media and marketing: A systematic review. *J. Bus. Res.* **2021**, *124*, 329–341. [CrossRef]
41. Visentin, M.; Pizzi, G.; Pichierri, M. Fake News, Real Problems for Brands: The Impact of Content Truthfulness and Source Credibility on consumers’ Behavioral Intentions toward the Advertised Brands. *J. Interact. Mark.* **2019**, *45*, 99–112. [CrossRef]
42. Di Domenico, G.; Visentin, M. Fake news or true lies? Reflections about problematic contents in marketing. *Int. J. Mark. Res.* **2020**, Forthcoming. [CrossRef]
43. Bakir, V.; McStay, A. Fake News and The Economy of Emotions. *Digit. J.* **2018**, *6*, 154–175. [CrossRef]
44. Sindermann, C.; Cooper, A.; Montag, C. A short review on susceptibility to falling for fake political news. *Curr. Opin. Psychol.* **2020**, *36*, 44–48. Cyberpsychology. [CrossRef] [PubMed]
45. Scardigno, R.; Mininni, G. The Rhetoric Side of Fake News: A New Weapon for Anti-Politics? *World Future* **2020**, *76*, 81–101. [CrossRef]
46. Brun, I. *National Security in the Era of Post-Truth and Fake News*; Institute for National Security Studies: Tel Aviv, Israel, 2020.
47. Belova, G.; Georgieva, G. Fake News as a Threat to National Security. *Int. Conf. Knowl. Based Organ.* **2018**, *24*, 19–22. [CrossRef]
48. Vasu, N.; Ang, B.; Teo, T.A.; Jayakumar, S.; Faizal, M.; Ahuja, J. *Fake News: National Security in the Post-Truth Era*; Technical Report; S. Rajaratnam School of International Studies, Nanyang Technological University: Singapore, 2018.
49. Verrall, N.; Mason, D. The Taming of the Shrewd. How Can the Military Tackle Sophistry, ‘Fake’ News and Post-Truth in the Digital Age? *RUSI J.* **2018**, *163*, 20–28. [CrossRef]
50. Gallacher, J.D.; Barash, V.; Howard, P.N.; Kelly, J. *Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans*; Data Memo 2017.9; Project on Computational Propaganda; Oxford Internet Institute, University of Oxford: Oxford, UK, 2017.
51. Kwanda, F.A.; Lin, T.T.C. Fake news practices in Indonesian newsrooms during and after the Palu earthquake: A hierarchy-of-influences approach. *Inf. Commun. Soc.* **2020**, *23*, 849–866. [CrossRef]
52. Hunt, K.; Agarwal, P.; Aziz, R.A.; Zhuang, D.J. Fighting Fake News during Disasters. *OR/MS Today* **2020**, *47*, 34–39. [CrossRef]
53. Sawano, T.; Ozaki, A.; Hori, A.; Tsubokura, M. Combating ‘fake news’ and social stigma after the Fukushima Daiichi Nuclear Power Plant incident—The importance of accurate longitudinal clinical data. *QJM Int. J. Med.* **2019**, *112*, 479–481. [CrossRef]

54. Naeem, S.B.; Bhatti, R.; Khan, A. An exploration of how fake news is taking over social media and putting public health at risk. *Health Inf. Libr. J.* **2020**. [[CrossRef](#)] [[PubMed](#)]
55. Anoop, K.; Deepak, P.; Lajish, V.L. Emotion Cognizance Improves Health Fake News Identification. In Proceedings of the 24th Symposium on International Database Engineering & Applications, Seoul, Korea, 12–14 August 2020; Association for Computing Machinery: New York, NY, USA, 2020; IDEAS '20. [[CrossRef](#)]
56. Dai, E.; Sun, Y.; Wang, S. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. In Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM 2020, Atlanta, GA, USA, 8–11 June 2020; Choudhury, M.D., Chunara, R., Culotta, A., Welles, B.F., Eds.; AAAI Press: Palo Alto, CA, USA, 2020; pp. 853–862.
57. Mesquita, C.T.; Oliveira, A.; Seixas, F.L.; Paes, A. Infodemia, Fake News and Medicine: Science and The Quest for Truth. *Int. J. Cardiovasc. Sci.* **2020**, *33*, 203–205. [[CrossRef](#)]
58. Hansen, P.R.; Schmidtblaicher, M. A Dynamic Model of Vaccine Compliance: How Fake News Undermined the Danish HPV Vaccine Program. *J. Bus. Econ. Stat.* **2021**, *39*, 259–271. [[CrossRef](#)]
59. Loomba, S.; de Figueiredo, A.; Piatek, S.J.; de Graaf, K.; Larson, H.J. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* **2021**. [[CrossRef](#)]
60. The Lancet Infectious Diseases. The COVID-19 infodemic. *Lancet Infect. Dis.* **2020**, *20*, 875. [[CrossRef](#)]
61. Ceron, W.; de Lima-Santos, M.F.; Quiles, M.G. Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online Soc. Netw. Media* **2021**, *21*, 100116. [[CrossRef](#)]
62. Englmeier, K. The Role of Text Mining in Mitigating the Threats from Fake News and Misinformation in Times of Corona. *Procedia Comput. Sci.* **2021**, *181*, 149–156. [[CrossRef](#)]
63. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [[CrossRef](#)] [[PubMed](#)]
64. Rubin, V.L.; Conroy, N. Discerning truth from deception: Human judgments and automation efforts. *First Monday* **2012**, *17*. [[CrossRef](#)]
65. Alonso García, S.; Gómez García, G.; Sanz Prieto, M.; Moreno Guerrero, A.J.; Rodríguez Jiménez, C. The Impact of Term Fake News on the Scientific Community. Scientific Performance and Mapping in Web of Science. *Soc. Sci.* **2020**, *9*, 73. [[CrossRef](#)]
66. De keersmaecker, J.; Roets, A. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* **2017**, *65*, 107–110. [[CrossRef](#)]
67. Tang, J.; Chang, Y.; Liu, H. Mining social media with social theories: A survey. *SIGKDD Explor.* **2013**, *15*, 20–29. [[CrossRef](#)]
68. Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. A Stylometric Inquiry into Hyperpartisan and Fake News. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018; Volume 1: Long Papers; Gurevych, I., Miyao, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 231–240. [[CrossRef](#)]
69. Parikh, S.B.; Atrey, P.K. Media-Rich Fake News Detection: A Survey. In Proceedings of the IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018, Miami, FL, USA, 10–12 April 2018; IEEE: Piscataway, NJ, USA, 2018, pp. 436–441. [[CrossRef](#)]
70. Wu, Y.; Agarwal, P.K.; Li, C.; Yang, J.; Yu, C. Toward Computational Fact-Checking. *Proc. VLDB Endow.* **2014**, *7*, 589–600. [[CrossRef](#)]
71. Ciampaglia, G.L.; Shiralkar, P.; Rocha, L.M.; Bollen, J.; Menczer, F.; Flammini, A. Computational Fact Checking from Knowledge Networks. *PLoS ONE* **2015**, *10*, e0128193. [[CrossRef](#)]
72. Magdy, A.; Wanas, N.M. Web-based statistical fact checking of textual documents. In Proceedings of the 2nd international workshop on Search and mining user-generated contents, SMUC@CIKM 2010, Toronto, ON, Canada, 30 October 2010; Cortizo, J.C., Carrero, F.M., Cantador, I., Jiménez, J.A.T., Rosso, P., Eds.; ACM: New York, NY, USA, 2010; pp. 103–110. [[CrossRef](#)]
73. Lease, M. Fact Checking and Information Retrieval. In Proceedings of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems, Bertinoro, Italy, 28–31 August 2018; Alonso, O., Silvello, G., Eds.; CEUR-WS.org, 2018; Volume 2167, CEUR Workshop Proceedings, pp. 97–98.
74. Hardalov, M.; Arora, A.; Nakov, P.; Augenstein, I. A Survey on Stance Detection for Mis- and Disinformation Identification. *arXiv* **2021**, arXiv:2103.00242.
75. Lillie, A.E.; Middelboe, E.R. Fake News Detection using Stance Classification: A Survey. *arXiv* **2019**, arXiv:1907.00181.
76. Vosoughi, S.; Mohsenvand, M.N.; Roy, D. Rumor Gauge: Predicting the Veracity of Rumors on Twitter. *ACM Trans. Knowl. Discov. Data* **2017**, *11*, 50:1–50:36. [[CrossRef](#)]
77. Li, Y.; Gao, J.; Meng, C.; Li, Q.; Su, L.; Zhao, B.; Fan, W.; Han, J. A Survey on Truth Discovery. *SIGKDD Explor.* **2015**, *17*, 1–16. [[CrossRef](#)]
78. Ott, M.; Cardie, C.; Hancock, J.T. Negative Deceptive Opinion Spam. In Proceedings of the Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, Westin Peachtree Plaza Hotel, Atlanta, GA, USA, 9–14 June 2013; Vanderwende, L., Daumé, H., III, Kirchhoff, K., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 497–501.
79. Peng, Q.; Zhong, M. Detecting Spam Review through Sentiment Analysis. *J. Softw.* **2014**, *9*, 2065–2072. [[CrossRef](#)]
80. Wu, Y.; Ngai, E.W.T.; Wu, P.; Wu, C. Fake online reviews: Literature review, synthesis, and directions for future research. *Decis. Support Syst.* **2020**, *132*, 113280. [[CrossRef](#)]
81. Elmogy, A.M.; Tariq, U.; Mohammed, A.; Ibrahim, A. Fake Reviews Detection using Supervised Machine Learning. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*. [[CrossRef](#)]

82. Latah, M. Detection of malicious social bots: A survey and a refined taxonomy. *Expert Syst. Appl.* **2020**, *151*, 113383. [[CrossRef](#)]
83. Ferrara, E.; Varol, O.; Davis, C.A.; Menczer, F.; Flammini, A. The rise of social bots. *Commun. ACM* **2016**, *59*, 96–104. [[CrossRef](#)]
84. Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, 28 March–1 April 2011; Srinivasan, S., Ramamritham, K., Kumar, A., Ravindra, M.P., Bertino, E., Kumar, R., Eds.; ACM: New York, NY, USA, 2011; pp. 675–684. [[CrossRef](#)]
85. Viviani, M.; Pasi, G. Credibility in social media: Opinions, news, and health information—A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*. [[CrossRef](#)]
86. Schmidt, A.; Wiegand, M. A Survey on Hate Speech Detection using Natural Language Processing. In Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, SocialNLP@EACL 2017, Valencia, Spain, 3 April 2017; Ku, L., Li, C., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 1–10. [[CrossRef](#)]
87. Fortuna, P.; Nunes, S. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.* **2018**, *51*, 85:1–85:30. [[CrossRef](#)]
88. Alrehili, A. Automatic Hate Speech Detection on Social Media: A Brief Survey. In Proceedings of the 16th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2019, Abu Dhabi, United Arab Emirates, 3–7 November 2019; IEEE Computer Society: Washington, DC, USA, 2019; pp. 1–6. [[CrossRef](#)]
89. Rubin, V.L.; Chen, Y.; Conroy, N.J. Deception detection for news: Three types of fakes. In Proceedings of the Information Science with Impact: Research in and for the Community—78th ASIS&T Annual Meeting, ASIST 2015, St. Louis, MO, USA, 6–10 October 2015; Wiley: Hoboken, NJ, USA, 2015; Volume 52, pp. 1–4. [[CrossRef](#)]
90. Ott, M.; Choi, Y.; Cardie, C.; Hancock, J.T. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, Portland, OR, USA, 19–24 June 2011; Lin, D., Matsumoto, Y., Mihalcea, R., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2011; pp. 309–319.
91. Ferreira, W.; Vlachos, A. Emergent: A novel data-set for stance classification. In Proceedings of the NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; Knight, K., Nenkova, A., Rambow, O., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 1163–1168. [[CrossRef](#)]
92. Zubiaga, A.; Liakata, M.; Procter, R.; Wong Sak Hoi, G.; Tolmie, P. Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. *PLoS ONE* **2016**, *11*, e0150989. [[CrossRef](#)] [[PubMed](#)]
93. Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.; Cha, M. Detecting Rumors from Microblogs with Recurrent Neural Networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016; Kambhampati, S., Ed.; IJCAI/AAAI Press: Palo Alto, CA, USA, 2016; pp. 3818–3824.
94. Tacchini, E.; Ballarin, G.; Vedova, M.L.D.; Moret, S.; de Alfaro, L. Some Like it Hoax: Automated Fake News Detection in Social Networks. *arXiv* **2017**, arXiv:1704.07506.
95. Pérez-Rosas, V.; Mihalcea, R. Experiments in Open Domain Deception Detection. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, 17–21 September 2015; Màrquez, L., Callison-Burch, C., Su, J., Pighin, D., Marton, Y., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2015; pp. 1120–1125. [[CrossRef](#)]
96. Vlachos, A.; Riedel, S. Fact Checking: Task definition and dataset construction. In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, Baltimore, MD, USA, 26 June 2014; pp. 18–22. [[CrossRef](#)]
97. Santia, G.C.; Williams, J.R. BuzzFace: A News Veracity Dataset with Facebook User Commentary and Egos. In Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, CA, USA, 25–28 June 2018; AAAI Press: Palo Alto, CA, USA, 2018; pp. 531–540.
98. Wang, W.Y. “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, BC, Canada, 30 July–4 August 2017; Volume 2: Short Papers; Barzilay, R., Kan, M., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 422–426. [[CrossRef](#)]
99. Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017; Palmer, M., Hwa, R., Riedel, S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 2931–2937. [[CrossRef](#)]
100. Khan, J.Y.; Khondaker, M.T.I.; Iqbal, A.; Afroz, S. A Benchmark Study on Machine Learning Methods for Fake News Detection. *arXiv* **2019**, arXiv:1905.04749.
101. Mitra, T.; Gilbert, E. CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations. In Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, 26–29 May 2015; Cha, M., Mascolo, C., Sandvig, C., Eds.; AAAI Press: Palo Alto, CA, USA, 2015; pp. 258–267.
102. Popat, K.; Mukherjee, S.; Strötgen, J.; Weikum, G. Credibility Assessment of Textual Claims on the Web. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; ACM: New York, NY, USA, 2016; CIKM '16, pp. 2173–2178. [[CrossRef](#)]

103. Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; Mittal, A. FEVER: A Large-scale Dataset for Fact Extraction and VERification. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics: New Orleans, LA, USA, 2018; pp. 809–819. [[CrossRef](#)]
104. Monteiro, R.A.; Santos, R.L.S.; Pardo, T.A.S.; de Almeida, T.A.; Ruiz, E.E.S.; Vale, O.A. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In Proceedings of the Computational Processing of the Portuguese Language-13th International Conference, PROPOR 2018, Canela, Brazil, 24–26 September 2018; Villavicencio, A., Moreira, V.P., Abad, A., de Medeiros Caseli, H., Gamallo, P., Ramisch, C., Oliveira, H.G., Paetzold, G.H., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11122, *Lecture Notes in Computer Science*, pp. 324–334. [[CrossRef](#)]
105. Silva, R.M.; Santos, R.L.; Almeida, T.A.; Pardo, T.A. Towards automatically filtering fake news in Portuguese. *Expert Syst. Appl.* **2020**, *146*, 113199. [[CrossRef](#)]
106. Pathak, A.; Srihari, R.K. BREAKING! Presenting Fake News Corpus for Automated Fact Checking. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 28 July–2 August 2019; Volume 2: Student Research Workshop; Alva-Manchego, F.E., Choi, E., Khashabi, D., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 357–362. [[CrossRef](#)]
107. Nørregaard, J.; Horne, B.D.; Adali, S. NELA-GT-2018: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles. In Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019, Munich, Germany, 11–14 June 2019; Pfeffer, J., Budak, C., Lin, Y., Morstatter, F., Eds.; AAAI Press: Palo Alto, CA, USA, 2019; pp. 630–638.
108. Gruppi, M.; Horne, B.D.; Adali, S. NELA-GT-2019: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. *arXiv* **2020**, arXiv:2003.08444.
109. Salem, F.K.A.; Feel, R.A.; Elbassuoni, S.; Jaber, M.; Farah, M. FA-KES: A Fake News Dataset around the Syrian War. In Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019, Munich, Germany, 11–14 June 2019; Pfeffer, J., Budak, C., Lin, Y., Morstatter, F., Eds.; AAAI Press: Palo Alto, CA, USA, 2019; pp. 573–582.
110. Posadas-Durán, J.P.; Gómez-Adorno, H.; Sidorov, G.; Escobar, J.J.M. Detection of fake news in a new corpus for the Spanish language. *J. Intell. Fuzzy Syst.* **2019**, *36*, 4869–4876. [[CrossRef](#)]
111. Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; Liu, H. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data* **2020**, *8*, 171–188. [[CrossRef](#)] [[PubMed](#)]
112. Zhang, X.; Cao, J.; Li, X.; Sheng, Q.; Zhong, L.; Shu, K. Mining Dual Emotion for Fake News Detection. In *The Web Conference 2021, Proceedings of The World Wide Web Conference WWW 2021*; ACM: New York, NY, USA, 2021. [[CrossRef](#)]
113. Hanselowski, A.; PVS, A.; Schiller, B.; Caspelherr, F.; Chaudhuri, D.; Meyer, C.M.; Gurevych, I. A Retrospective Analysis of the Fake News Challenge Stance-Detection Task. In Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, NM, USA, 20–26 August 2018; Bender, E.M., Derczynski, L., Isabelle, P., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 1859–1874.
114. Bevendorff, J.; Ghanem, B.; Giachanou, A.; Kestemont, M.; Manjavacas, E.; Markov, I.; Mayerl, M.; Potthast, M.; Rangel, F.; Rosso, P.; et al. Overview of PAN 2020: Authorship Verification, Celebrity Profiling, Profiling Fake News Spreaders on Twitter, and Style Change Detection. In Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020), Thessaloniki, Greece, 22–25 September 2020; Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Névél, A., Cappellato, L., Ferro, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2020.
115. Amjad, M.; Sidorov, G.; Zhila, A.; Gelbukh, A.; Rosso, P. UrduFake@FIRE2020: Shared Track on Fake News Identification in Urdu. In *Forum for Information Retrieval Evaluation*; Association for Computing Machinery: New York, NY, USA, 2020; FIRE 2020, pp. 37–40. [[CrossRef](#)]
116. Patwa, P.; Bhardwaj, M.; Guptha, V.; Kumari, G.; Sharma, S.; PYKL, S.; Das, A.; Ekbal, A.; Akhtar, S.; Chakraborty, T. Overview of CONSTRAINT 2021 Shared Tasks: Detecting English COVID-19 Fake News and Hindi Hostile Posts. In *Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT)*; Springer: Berlin/Heidelberg, Germany, 2021.
117. Patwa, P.; Sharma, S.; Srinivas, P.; Guptha, V.; Kumari, G.; Akhtar, M.S.; Ekbal, A.; Das, A.; Chakraborty, T. Fighting an Infodemic: COVID-19 Fake News Dataset. *arXiv* **2020**, arXiv:2011.03327.
118. Bhardwaj, M.; Akhtar, M.S.; Ekbal, A.; Das, A.; Chakraborty, T. Hostility Detection Dataset in Hindi. *arXiv* **2020**, arXiv:2011.03588.
119. Ling, C.X.; Huang, J.; Zhang, H. AUC: A Statistically Consistent and more Discriminating Measure than Accuracy. In Proceedings of the IJCAI-03, Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 9–15 August 2003; Gottlob, G., Walsh, T., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 2003; pp. 519–526.
120. Wiebe, J.; Bruce, R.F.; O'Hara, T.P. Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. In Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, University of Maryland, College Park, MD, USA, 20–26 June 1999; Dale, R., Church, K.W., Eds.; ACL: Stroudsburg, PA, USA, 1999; pp. 246–253. [[CrossRef](#)]
121. Turney, P.D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 6–12 July 2002; ACL: Stroudsburg, PA, USA, 2002; pp. 417–424. [[CrossRef](#)]

122. Esuli, A.; Sebastiani, F. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC 2006, Genoa, Italy, 22–28 May 2006; Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J., Tapias, D., Eds.; European Language Resources Association (ELRA): Paris, France, 2006; pp. 417–422.
123. Miller, G.A. WordNet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41. [[CrossRef](#)]
124. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.D.; Stede, M. Lexicon-Based Methods for Sentiment Analysis. *Comput. Linguist.* **2011**, *37*, 267–307. [[CrossRef](#)]
125. Brooke, J.; Tofiloski, M.; Taboada, M. Cross-Linguistic Sentiment Analysis: From English to Spanish. In Proceedings of the Recent Advances in Natural Language Processing, RANLP 2009, Borovets, Bulgaria, 14–16 September 2009; Angelova, G., Bontcheva, K., Mitkov, R., Nicolov, N., Nikolov, N., Eds.; RANLP 2009 Organising Committee/ACL: Stroudsburg, PA, USA, 2009; pp. 50–54.
126. Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; Kappas, A. Sentiment in short strength detection informal text. *J. Assoc. Inf. Sci. Technol.* **2010**, *61*, 2544–2558. [[CrossRef](#)]
127. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, MI, USA, 1–4 June 2014; Adar, E., Resnick, P., Choudhury, M.D., Hogan, B., Oh, A.H., Eds.; The AAAI Press: Palo Alto, CA, USA, 2014.
128. Vilares, D.; Gómez-Rodríguez, C.; Alonso, M.A. Universal, unsupervised (rule-based), uncovered sentiment analysis. *Knowl. Based Syst.* **2017**, *118*, 45–55. [[CrossRef](#)]
129. Cambria, E.; Olsher, D.; Rajagopal, D. SenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014; Brodley, C.E., Stone, P., Eds.; AAAI Press: Palo Alto, CA, USA, 2014; pp. 1515–1521.
130. Cambria, E.; Li, Y.; Xing, F.Z.; Poria, S.; Kwok, K. SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis. In Proceedings of the CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, 19–23 October 2020; d'Aquin, M., Dietze, S., Hauff, C., Curry, E.; Cudré-Mauroux, P., Eds.; ACM: New York, NY, USA, 2020; pp. 105–114. [[CrossRef](#)]
131. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, Philadelphia, PA, USA, 6–7 July 2002; pp. 79–86. [[CrossRef](#)]
132. Mohammad, S.; Kiritchenko, S.; Zhu, X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2013, Atlanta, GA, USA, 14–15 June 2013; Diab, M.T., Baldwin, T., Baroni, M., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2013; pp. 321–327.
133. Agarwal, A.; Xie, B.; Vovsha, I.; Rambow, O.; Passonneau, R. Sentiment Analysis of Twitter Data. In *Proceedings of the Workshop on Languages in Social Media*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; LSM '11, pp. 30–38.
134. Joshi, M.; Rosé, C.P. Generalizing Dependency Features for Opinion Mining. In Proceedings of the ACL 2009, 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, 2–7 August 2009; Short Papers; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2009; pp. 313–316.
135. Vilares, D.; Alonso, M.A.; Gómez-Rodríguez, C. On the usefulness of lexical and syntactic processing in polarity classification of Twitter messages. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1799–1816. [[CrossRef](#)]
136. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, Baltimore, MD, USA, 22–27 June 2014; Volume 1: Long Papers; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2014; pp. 655–665. [[CrossRef](#)]
137. dos Santos, C.N.; Gatti, M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In Proceedings of the COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, Dublin, Ireland, 23–29 August 2014; Hajic, J., Tsujii, J., Eds.; ACL: Stroudsburg, PA, USA, 2014; pp. 69–78.
138. Severyn, A.; Moschitti, A. Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; Baeza-Yates, R., Lalmas, M., Moffat, A., Ribeiro-Neto, B.A., Eds.; ACM: New York, NY, USA, 2015; pp. 959–962. [[CrossRef](#)]
139. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, Grand Hyatt Seattle, Seattle, WA, USA, 18–21 October 2013; A meeting of SIGDAT, a Special Interest Group of the ACL; ACL: Stroudsburg, PA, USA, 2013; pp. 1631–1642.
140. Tai, K.S.; Socher, R.; Manning, C.D. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, Beijing, China, 26–31 July 2015; Volume 1: Long Papers; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2015; pp. 1556–1566. [[CrossRef](#)]

141. Radford, A.; Józefowicz, R.; Sutskever, I. Learning to Generate Reviews and Discovering Sentiment. *arXiv* **2017**, arXiv:1704.01444.
142. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; Volume 1 (Long and Short Papers); Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186. [[CrossRef](#)]
143. Dang, N.C.; Moreno-García, M.N.; De la Prieta, F. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics* **2020**, *9*, 483. [[CrossRef](#)]
144. Vilares, D.; Alonso, M.A.; Gómez-Rodríguez, C. A syntactic approach for opinion mining on Spanish reviews. *Nat. Lang. Eng.* **2015**, *21*, 139–163. [[CrossRef](#)]
145. Vilares, D.; Alonso, M.A.; Gómez-Rodríguez, C. Supervised sentiment analysis in multilingual environments. *Inf. Process. Manag.* **2017**, *53*, 595–607. [[CrossRef](#)]
146. Hamidian, S.; Diab, M.T. Rumor Identification and Belief Investigation on Twitter. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@NAACL-HLT 2016, San Diego, CA, USA, 16 June 2016; Balahur, A., der Goot, E.V., Vossen, P., Montoyo, A., Eds.; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2016; pp. 3–8. [[CrossRef](#)]
147. Meire, M.; Ballings, M.; den Poel, D.V. The added value of auxiliary data in sentiment analysis of Facebook posts. *Decis. Support Syst.* **2016**, *89*, 98–112. [[CrossRef](#)]
148. AbdelFattah, M.; Galal, D.; Hassan, N.; Elzanfaly, D.; Tallent, G. A Sentiment Analysis Tool for Determining the Promotional Success of Fashion Images on Instagram. *Int. J. Interact. Mob. Technol.* **2017**, *11*, 66–73. [[CrossRef](#)]
149. Hu, X.; Tang, J.; Gao, H.; Liu, H. Social Spammer Detection with Sentiment Information. In Proceedings of the 2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, 14–17 December 2014; Kumar, R., Toivonen, H., Pei, J., Huang, J.Z., Wu, X., Eds.; IEEE Computer Society: Washington, DC, USA, 2014; pp. 180–189. [[CrossRef](#)]
150. Rubin, V.; Conroy, N.; Chen, Y.; Cornwell, S. Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In Proceedings of the Second Workshop on Computational Approaches to Deception Detection; Association for Computational Linguistics: San Diego, CA, USA, 2016; pp. 7–17. [[CrossRef](#)]
151. Zhang, S.; Zhang, X.; Chan, J.; Rosso, P. Irony detection via sentiment-based transfer learning. *Inf. Process. Manag.* **2019**, *56*, 1633–1644. [[CrossRef](#)]
152. Arcila-Calderón, C.; Blanco-Herrero, D.; Frías-Vázquez, M.; Seoane-Pérez, F. Refugees Welcome? Online Hate Speech and Sentiments in Twitter in Spain during the Reception of the Boat Aquarius. *Sustainability* **2021**, *13*, 2728. [[CrossRef](#)]
153. Li, S.; Li, G.; Law, R.; Paradies, Y. Racism in tourism reviews. *Tour. Manag.* **2020**, *80*, 104100. [[CrossRef](#)]
154. Jha, A.; Mamidi, R. When does a compliment become sexist? Analysis and classification of ambivalent sexism using twitter data. In Proceedings of the Second Workshop on NLP and Computational Social Science, NLP+CSS@ACL 2017, Vancouver, BC, Canada, 3 August 2017; Hovy, D., Volkova, S., Bamman, D., Jurgens, D., O'Connor, B., Tsur, O., Dogruöz, A.S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 7–16. [[CrossRef](#)]
155. Aguilar, S.J.; Baek, C. Sexual harassment in academe is underreported, especially by students in the life and physical sciences. *PLoS ONE* **2020**, *15*, e0230312. [[CrossRef](#)] [[PubMed](#)]
156. Dessì, D.; Recupero, D.R.; Sack, H. An Assessment of Deep Learning Models and Word Embeddings for Toxicity Detection within Online Textual Comments. *Electronics* **2021**, *10*, 779. [[CrossRef](#)]
157. Vilares, D.; Hermo, M.; Alonso, M.A.; Gómez-Rodríguez, C.; Vilares, J. LyS at CLEF RepLab 2014: Creating the State of the Art in Author Influence Ranking and Reputation Classification on Twitter. In Proceedings of the Working Notes for CLEF 2014 Conference, Sheffield, UK, 15–18 September 2014; Cappellato, L., Ferro, N., Halvey, M., Kraaij, W., Eds.; CEUR-WS.org: Aachen, Germany, 2014; Volume 1180, CEUR Workshop Proceedings, pp. 1468–1478.
158. Bamakan, S.M.H.; Nurgaliev, I.; Qu, Q. Opinion leader detection: A methodological review. *Expert Syst. Appl.* **2019**, *115*, 200–222. [[CrossRef](#)]
159. Vilares, D.; Thelwall, M.; Alonso, M.A. The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *J. Inf. Sci.* **2015**, *41*, 799–813. [[CrossRef](#)]
160. Alonso, M.A.; Vilares, D. A review on political analysis and social media. *Proces. Leng. Nat.* **2016**, *56*, 13–24.
161. Jang, J.-S.; Lee, B.-I.; Choi, C.-H.; Kim, J.-H.; Seo, D.-M.; Cho, W.-S. Understanding pending issue of society and sentiment analysis using social media. In Proceedings of the 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, Austria, 5–8 July 2016; pp. 981–986. [[CrossRef](#)]
162. Etter, M.; Colleoni, E.; Illia, L.; Meggiorin, K.; D'Eugenio, A. Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. *Bus. Soc.* **2018**, *57*, 60–97. [[CrossRef](#)]
163. Azar, P.D.; Lo, A.W. The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds. *J. Portf. Manag.* **2016**, *42*, 123–134. [[CrossRef](#)]
164. Chen, H.; De, P.; Hu, Y.J.; Hwang, B.H. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Rev. Financ. Stud.* **2014**, *27*, 1367–1403. [[CrossRef](#)]
165. Sharma, S.; Jain, A. Role of sentiment analysis in social media security and analytics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*. [[CrossRef](#)]

166. Zunic, A.; Corcoran, P.; Spasic, I. Sentiment Analysis in Health and Well-Being: Systematic Review. *JMIR Med Inform.* **2020**, *8*, e16023. [[CrossRef](#)]
167. Alamoodi, A.H.; Zaidan, B.B.; Zaidan, A.A.; Albahri, O.S.; Mohammed, K.I.; Malik, R.Q.; Almahdi, E.M.; Chyad, M.A.; Tareq, Z.; Albahri, A.S.; et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Syst. Appl.* **2021**, *167*, 114155. [[CrossRef](#)]
168. Diakopoulos, N.; Naaman, M.; Kivran-Swaine, F. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In Proceedings of the 5th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2010, Salt Lake City, UT, USA, 24–29 October 2010; Part of VisWeek 2010; IEEE Computer Society: Washington, DC, USA, 2010; pp. 115–122. [[CrossRef](#)]
169. AlRubaian, M.A.; Al-Qurishi, M.; Al-Rakhami, M.; Rahman, S.M.M.; Alamri, A. A Multistage Credibility Analysis Model for Microblogs. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, Paris, France, 25–28 August 2015; Pei, J., Silvestri, F., Tang, J., Eds.; ACM: New York, NY, USA, 2015; pp. 1434–1440. [[CrossRef](#)]
170. Chatterjee, R.; Agarwal, S. Twitter truths: Authenticating analysis of information credibility. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 2352–2357.
171. Dey, A.; Rafi, R.Z.; Hasan Parash, S.; Arko, S.K.; Chakrabarty, A. Fake News Pattern Recognition using Linguistic Analysis. In Proceedings of the 2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR), Kitakyushu, Japan, 25–29 June 2018; pp. 305–309. [[CrossRef](#)]
172. Bhutani, B.; Rastogi, N.; Sehgal, P.; Purwar, A. Fake News Detection Using Sentiment Analysis. In Proceedings of the 2019 Twelfth International Conference on Contemporary Computing, IC3 2019, Noida, India, 8–10 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5. [[CrossRef](#)]
173. Ajao, O.; Bhowmik, D.; Zargari, S. Sentiment Aware Fake News Detection on Online Social Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2507–2511. [[CrossRef](#)]
174. Zubiaga, A.; Liakata, M.; Procter, R. Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media. *arXiv* **2016**, arXiv:1610.07363.
175. Cui, L.; Wang, S.; Lee, D. SAME: Sentiment-aware multi-modal embedding for detecting fake news. In Proceedings of the ASONAM '19: International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada, 27–30 August 2019; Spezzano, F., Chen, W., Xiao, X., Eds.; ACM: New York, NY, USA, 2019; pp. 41–48. [[CrossRef](#)]
176. Vicario, M.D.; Quattrociocchi, W.; Scala, A.; Zollo, F. Polarization and Fake News: Early Warning of Potential Misinformation Targets. *ACM Trans. Web* **2019**, *13*. [[CrossRef](#)]
177. Ross, J.; Thirunarayan, K. Features for Ranking Tweets Based on Credibility and Newsworthiness. In Proceedings of the 2016 International Conference on Collaboration Technologies and Systems, CTS 2016, Orlando, FL, USA, 31 October–4 November 2016; Smari, W.W., Natarian, J., Eds.; IEEE Computer Society: Washington, DC, USA, 2016; pp. 18–25. [[CrossRef](#)]
178. Nakashole, N.; Mitchell, T.M. Language-Aware Truth Assessment of Fact Candidates. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, Baltimore, MD, USA, 22–27 June 2014; Volume 1: Long Papers; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2014; pp. 1009–1019. [[CrossRef](#)]
179. Popat, K.; Mukherjee, S.; Strötgen, J.; Weikum, G. Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; International World Wide Web Conferences Steering Committee: Republic and Canton of Geneva, Switzerland, 2017; WWW '17 Companion, pp. 1003–1012. [[CrossRef](#)]
180. Hassan, N.; Arslan, F.; Li, C.; Tremayne, M. Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; ACM: New York, NY, USA, 2017; pp. 1803–1812. [[CrossRef](#)]
181. Alonso, M.A.; Gómez-Rodríguez, C.; Vilares, J. On the Use of Parsing for Named Entity Recognition. *Appl. Sci.* **2021**, *11*, 1090. [[CrossRef](#)]
182. Varol, O.; Ferrara, E.; Menczer, F.; Flammini, A. Early detection of promoted campaigns on social media. *EPJ Data Sci.* **2017**, *6*, 13. [[CrossRef](#)]
183. Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P.S. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv* **2018**, arXiv:1806.00749.
184. Reis, J.C.S.; Correia, A.; Murai, F.; Veloso, A.; Benevenuto, F.; Cambria, E. Supervised Learning for Fake News Detection. *IEEE Intell. Syst.* **2019**, *34*, 76–81. [[CrossRef](#)]
185. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.J.; Hovy, E.H. Hierarchical Attention Networks for Document Classification. In Proceedings of the NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego CA, USA, 12–17 June 2016; Knight, K., Nenkova, A., Rambow, O., Eds.; The Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 1480–1489. [[CrossRef](#)]

186. Enayet, O.; El-Beltagy, S.R. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, BC, Canada, 3–4 August 2017; Bethard, S., Carpuat, M., Apidianaki, M., Mohammad, S.M., Cer, D.M., Jurgens, D., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 470–474. [[CrossRef](#)]
187. Guo, H.; Cao, J.; Zhang, Y.; Guo, J.; Li, J. Rumor Detection with Hierarchical Social Attention Network. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, 22–26 October 2018; Cuzzocrea, A., Allan, J., Paton, N.W., Srivastava, D., Agrawal, R., Broder, A.Z., Zaki, M.J., Candan, K.S., Labrinidis, A., Schuster, A., et al., Eds.; ACM: New York, NY, USA, 2018; pp. 943–951. [[CrossRef](#)]
188. Gorrell, G.; Aker, A.; Bontcheva, K.; Derczynski, L.; Kochkina, E.; Liakata, M.; Zubiaga, A. SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours. In Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, 6–7 June 2019; May, J., Shutova, E., Herbelot, A., Zhu, X., Apidianaki, M., Mohammad, S.M., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 845–854. [[CrossRef](#)]
189. Abonizio, H.Q.; de Morais, J.I.; Tavares, G.M.; Barbon Junior, S. Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features. *Future Internet* **2020**, *12*, 87. [[CrossRef](#)]
190. Guibon, G.; Ermakova, L.; Seffih, H.; Firsov, A.; Noé-Bienvenu, G.L. Fake News Detection with Satire. In *CICLing: International Conference on Computational Linguistics and Intelligent Text Processing*; CICLing: La Rochelle, France, 2019.
191. Li, Z.; Fan, Y.; Jiang, B.; Lei, T.; Liu, W. A survey on sentiment analysis and opinion mining for social multimedia. *Multim. Tools Appl.* **2019**, *78*, 6939–6967. [[CrossRef](#)]
192. Zhou, Z.; Guan, H.; Bhat, M.M.; Hsu, J. Fake News Detection via NLP is Vulnerable to Adversarial Attacks. In Proceedings of the 11th International Conference on Agents and Artificial Intelligence, ICAART 2019, Prague, Czech Republic, 19–21 February 2019; Rocha, A.P., Steels, L., van den Herik, H.J., Eds.; SciTePress: Setubal, Portugal, 2019; Volume 2, pp. 794–800. [[CrossRef](#)]
193. Do, H.H.; Prasad, P.W.C.; Maag, A.; Alsadoon, A. Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [[CrossRef](#)]
194. Han, J.; Zhang, Z.; Schuller, B.W. Adversarial Training in Affective Computing and Sentiment Analysis: Recent Advances and Perspectives [Review Article]. *IEEE Comput. Intell. Mag.* **2019**, *14*, 68–81. [[CrossRef](#)]
195. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [[CrossRef](#)] [[PubMed](#)]
196. Blodgett, S.L.; Barocas, S.; Daumé, H., III; Wallach, H.M. Language (Technology) is Power: A Critical Survey of “Bias” in NLP. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 5–10 July 2020; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 5454–5476. [[CrossRef](#)]
197. Bender, E.M.; Friedman, B. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Trans. Assoc. Comput. Linguist.* **2018**, *6*, 587–604. [[CrossRef](#)]
198. Kiritchenko, S.; Mohammad, S. Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, *SEM@NAACL-HLT 2018, New Orleans, LA, USA, 5–6 June 2018; Nissim, M., Berant, J., Lenci, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 43–53. [[CrossRef](#)]
199. van der Linden, S.; Panagopoulos, C.; Roozenbeek, J. You are fake news: Political bias in perceptions of fake news. *Media Cult. Soc.* **2020**, *42*, 460–470. [[CrossRef](#)]
200. Pennycook, G.; Rand, D.G. The Psychology of Fake News. *Trends Cogn. Sci.* **2021**, *25*, 388–402. [[CrossRef](#)] [[PubMed](#)]