



# Approaches to Identify Fake News: A Systematic Literature Review

Dylan de Beer and Machdel Matthee<sup>✉</sup> 

Department of Informatics, University of Pretoria, Pretoria 0001, South Africa  
machdel.matthee@up.ac.za

**Abstract.** With the widespread dissemination of information via digital media platforms, it is of utmost importance for individuals and societies to be able to judge the credibility of it. Fake news is not a recent concept, but it is a commonly occurring phenomenon in current times. The consequence of fake news can range from being merely annoying to influencing and misleading societies or even nations. A variety of approaches exist to identify fake news. By conducting a systematic literature review, we identify the main approaches currently available to identify fake news and how these approaches can be applied in different situations. Some approaches are illustrated with a relevant example as well as the challenges and the appropriate context in which the specific approach can be used.

**Keywords:** Fake news · Machine learning · Linguistics · Semantics · Syntax · Algorithms · Digital tools · Social media

## 1 Introduction

Paskin (2018: 254) defines fake news as “particular news articles that originate either on mainstream media (online or offline) or social media and have no factual basis, but are presented as facts and not satire”. The importance of combatting fake news is starkly illustrated during the current COVID-19 pandemic. Social networks are stepping up in using digital fake news detection tools and educating the public towards spotting fake news. At the time of writing, Facebook uses machine learning algorithms to identify false or sensational claims used in advertising for alternative cures, they place potential fake news articles lower in the news feed, and they provide users with tips on how to identify fake news themselves (Sparks and Frishberg 2020). Twitter ensures that searches on the virus result in credible articles and Instagram redirects anyone searching for information on the virus to a special message with credible information (Marr 2020).

These measures are possible because different approaches exist that assist the detection of fake news. For example, platforms based on machine learning use fake news from the biggest media outlets, to refine algorithms for identifying fake news (Macaulay 2018). Some approaches detect fake news by using metadata such as a comparison of release time of the article and timelines of spreading the article as well where the story spread (Macaulay 2018).

The purpose of this research paper is to, through a systematic literature review, categorize current approaches to contest the wide-ranging endemic of fake news.

## 2 The Evolution of Fake News and Fake News Detection

Fake news is not a new concept. Before the era of digital technology, it was spread through mainly yellow journalism with focus on sensational news such as crime, gossip, disasters and satirical news (Stein-Smith 2017). The prevalence of fake news relates to the availability of mass media digital tools (Schade 2019). Since anyone can publish articles via digital media platforms, online news articles include well researched pieces but also opinion-based arguments or simply false information (Burkhardt 2017). There is no custodian of credibility standards for information on these platforms making the spread of fake news possible. To make things worse, it is by no means straightforward telling the difference between real news and semi-true or false news (Pérez-Rosas et al. 2018).

The nature of social media makes it easy to spread fake news, as a user potentially sends fake news articles to friends, who then send it again to their friends and so on. Comments on fake news sometimes fuel its ‘credibility’ which can lead to rapid sharing resulting in further fake news (Albright 2017).

Social bots are also responsible for the spreading of fake news. Bots are sometimes used to target super-users by adding replies and mentions to posts. Humans are manipulated through these actions to share the fake news articles (Shao et al. 2018).

Clickbait is another tool encouraging the spread of fake news. Clickbait is an advertising tool used to get the attention of users. Sensational headlines or news are often used as clickbait that navigate the user to advertisements. More clicks on the advert means more money (Chen et al. 2015a).

Fortunately, tools have been developed for detecting fake news. For example, a tool has been developed to identify fake news that spreads through social media through examining lexical choices that appear in headlines and other intense language structures (Chen et al. 2015b). Another tool, developed to identify fake news on Twitter, has a component called the Twitter Crawler which collects and stores tweets in a database (Atodiresei et al. 2018). When a Twitter user wants to check the accuracy of the news found they can copy a link into this application after which the link will be processed for fake news detection. This process is built on an algorithm called the NER (Named Entity Recognition) (Atodiresei et al. 2018).

There are many available approaches to help the public to identify fake news and this paper aims to enhance understanding of these by categorizing these approaches as found in existing literature.

### 3 Research Method

#### 3.1 Research Objective

The purpose of this paper is to categorize approaches used to identify fake news. In order to do this, a systematic literature review was done. This section presents the search terms that were used, the selection criteria and the source selection.

#### 3.2 Search Terms

Specific search terms were used to enable the finding of relevant journal articles such as the following:

("what is fake news" OR "not genuine information" OR "counter fit news" OR "inaccurate report\*" OR "forged (NEAR/2) news" OR "mislead\* information" OR "false store\*" OR "untrustworthy information" OR "hokes" OR "doubtful information" OR "incorrect detail\*" OR "false news" OR "fake news" OR "false accusation\*")

AND ("digital tool\*" OR "digital approach" OR "automated tool\*" OR "approach\*" OR "programmed tool\*" OR "digital gadget\*" OR "digital device\*" OR "digital machan\*" OR "digital appliance\*" OR "digital gizmo" OR "IS gadget\*" OR "IS tool\*" OR "IS machine\*" OR "digital gear\*" OR "information device\*")

AND ("fake news detection" OR "approaches to identify fake news" OR "methods to identify fake news" OR "finding fake news" OR "ways to detect fake news").

#### 3.3 Selection Criteria

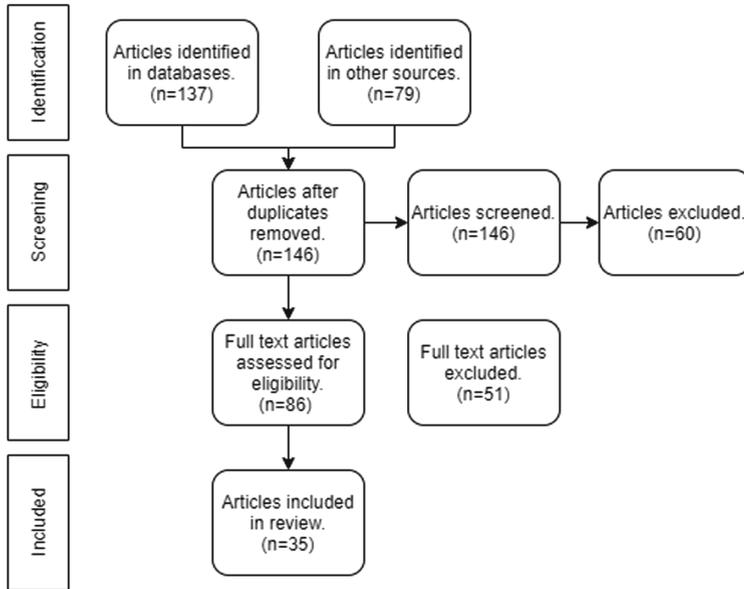
*Inclusion Criteria.* Studies that adhere to the following criteria: (1) studies published between 2008 and 2019; (2) studies found in English; (3) with main focus fake news on digital platforms; (4) articles that are published in IT journals or any technology related journal articles (e.g. computers in human behavior) as well as conference proceedings; (5) journal articles that are cited more than 10 times.

*Exclusion Criteria.* Studies that adhered to the following criteria: (1) studies not presented in journal articles (e.g. in the form of a slide show or overhead presentation); (2) studies published, not relating to technology or IT; (3) articles on fake news but not the identification of it.

The search terms were used to find relevant articles on ProQuest, ScienceDirect, EBSCOhost and Google Scholar (seen here as 'other sources').

#### 3.4 Flowchart of Search Process

Figure 1 below gives a flowchart of the search process: the identification of articles, the screening, the selection process and the number of the included articles.



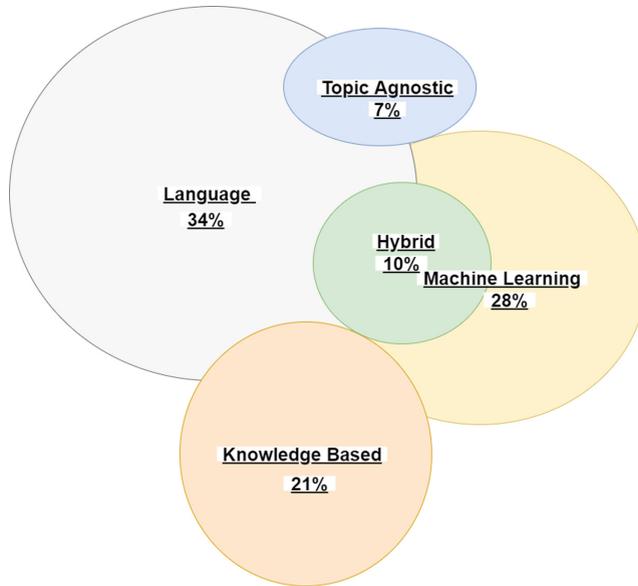
**Fig. 1.** A flowchart of the selection process

## 4 Findings

In this section of the article we list the categories of approaches that are used to identify fake news. We also discuss how the different approaches interlink with each other and how they can be used together to get a better result.

The following categories of approaches for fake news detection are proposed: (1) language approach, (2) topic-agnostic approach, (3) machine learning approach, (4) knowledge-based approach, (5) hybrid approach.

The five categories mentioned above are depicted in Fig. 2 below. Figure 2 shows the relationship between the different approaches. The sizes of the ellipses are proportional to the number of articles found (given as the percentage of total included articles) in the systematic literature review that refer to that approach.



**Fig. 2.** Categories of fake news detection approaches resulting from the systematic literature review

The approaches are discussed in depth below with some examples for illustration purposes.

#### 4.1 Language Approach

This approach focuses on the use of linguistics by a human or software program to detect fake news. Most of the people responsible for the spread of fake news have control over what their story is about, but they can often be exposed through the style of their language (Yang et al. 2018). The approach considers all the words in a sentence and letters in a word, how they are structured and how it fits together in a paragraph (Burkhardt 2017). The focus is therefore on grammar and syntax (Burkhardt 2017). There are currently three main methods that contribute to the language approach:

*Bag of Words (BOW):* In this approach, each word in a paragraph is considered of equal importance and as independent entities (Burkhardt 2017). Individual words frequencies are analysed to find signs of misinformation. These representations are also called n-grams (Thota et al. 2018). This will ultimately help to identify patterns of word use and by investigating these patterns, misleading information can be identified. The bag of words model is not as practical because context is not considered when text is converted into numerical representations and the position of a word is not always taken into consideration (Potthast et al. 2017).

*Semantic Analysis:* Chen et al. 2017b explain that truthfulness can be determined by comparing personal experience (e.g. restaurant review) with a profile on the topic

derived from similar articles. An honest writer will be more likely to make similar remarks about a topic than other truthful writers. Different compatibility scores are used in this approach.

*Deep Syntax*: The deep syntax method is carried out through Probability Context Free Grammars (Stahl 2018). The Probability Context Free Grammars executes deep syntax tasks through parse trees that make Context Free Grammar analysis possible. Probabilistic Context Free Grammar is an extension of Context Free Grammars (Zhou and Zafarani 2018). Sentences are converted into a set of rewritten rules and these rules are used to analyse various syntax structures. The syntax can be compared to known structures or patterns of lies and can ultimately lead to telling the difference between fake news and real news (Burkhardt 2017).

## 4.2 Topic-Agnostic Approach

This category of approaches detect fake news by not considering the content of articles but rather topic-agnostic features. The approach uses linguistic features and web markup capabilities to identify fake news (Castelo et al. 2019). Some examples of topic-agnostic features are 1) a large number of advertisements, 2) longer headlines with eye-catching phrases, 3) different text patterns from mainstream news to induce emotive responses 4) presence of an author name (Castelo et al. 2019; Horne and Adali 2017).

## 4.3 Machine Learning Approach

Machine learning algorithms can be used to identify fake news. This is achieved through using different types of training datasets to refine the algorithms. Datasets enables computer scientists to develop new machine learning approaches and techniques. Datasets are used to train the algorithms to identify fake news. How are these datasets created? One way is through crowdsourcing. Perez-Rosas et al. (2018) created a fake news data set by first collecting legitimate information on six different categories such as sports, business, entertainment, politics, technology and education (Pérez-Rosas et al. 2018). Crowdsourcing was then used and a task was set up which asked the workers to generate a false version of the news stories (Pérez-Rosas et al. 2018). Over 240 stories were collected and added to the fake news dataset.

A machine learning approach called the rumor identification framework has been developed that legitimizes signals of ambiguous posts so that a person can easily identify fake news (Sivasangari et al. 2018). The framework will alert people of posts that might be fake (Sivasangari et al. 2018). The framework is built to combat fake tweets on Twitter and focuses on four main areas; the metadata of tweets, the source of the tweet; the date and area of the tweet, where and when the tweet was developed (Sivasangari et al. 2018). By studying these four parts of the tweet the framework can be implemented to check the accuracy of the information and to separate the real from the fake (Sivasangari et al. 2018). Supporting this framework, the spread of gossip is collected to create datasets with the use of a Twitter Streaming API (Sivasangari et al. 2018).

Twitter has developed a possible solution to identify and prevent the spread of misleading information through fake accounts, likes and comments (Atodiresei et al. 2018) - the Twitter crawler, a machine learning approach works by collecting tweets and adding them to a database, making comparison between different tweets possible.

#### 4.4 Knowledge Based Approach

Recent studies argue for the integration of machine learning and knowledge engineering to detect fake news. The challenging problem with some of these fact checking methods is the speed at which fake news spreads on social media. Microblogging platforms such as Twitter causes small pieces of false information to spread very quickly to a large number of people (Qazvinian et al. 2011). The knowledge-based approach aims at using sources that are external to verify if the news is fake or real and to identify the news before the spread thereof becomes quicker. There are three main categories; (1) Expert Oriented Fact Checking, (2) Computational Oriented Fact Checking, (3) Crowd Sourcing Oriented Fact Checking (Ahmed et al. 2019).

*Expert Oriented Fact Checking.* With expert oriented fact checking it is necessary to analyze and examine data and documents carefully (Ahmed et al. 2019). Expert-oriented fact-checking requires professionals to evaluate the accuracy of the news manually through research and other studies on the specific claim. Fact checking is the process of assigning certainty to a specific element by comparing the accuracy of the text to another which has previously been fact checked (Vlachos and Riedel 2014).

*Computational Oriented Fact Checking.* The purpose of computational oriented fact checking is to administer users with an automated fact-checking process that is able to identify if a specific piece of news is true or false (Ahmed et al. 2019). An example of computational oriented fact checking is knowledge graphs and open web sources that are based on practical referencing to help distinguish between real and fake news (Ahmed et al. 2019). A recent tool called the ClaimBuster has been developed and is an example of how fact checking can automatically identify fake news (Hassan et al. 2017). This tool makes use of machine learning techniques combined with natural language processing and a variety of database queries. It analyses context on social media, interviews and speeches in real time to determine ‘facts’ and compares it with a repository that contains verified facts and delivers it to the reader (Hassan et al. 2017).

*Crowd Sourcing Oriented.* Crowdsourcing gives the opportunity for a group of people to make a collective decision through examining the accuracy of news (Pennycook and Rand 2019). The accuracy of the news is completely based on the wisdom of the crowd (Ahmed et al. 2019). Kiskkit is an example of a platform that can be used for crowdsourcing where the platform allows a group of people to evaluate pieces of a news article (Hassan et al. 2017). After one piece has been evaluated the crowd moves to the next piece for evaluation until the entire news article has been evaluated and the accuracy thereof has been determined by the wisdom of the crowd (Hassan et al. 2017).

## 4.5 Hybrid Approach

There are three generally agreed upon elements of fake news articles, the first element is the text of an article, second element is the response that the articles received and lastly the source used that motivate the news article (Ruchansky et al. 2017). A recent study has been conducted that proposes a hybrid model which helps to identify fake news on social media through using a combination of human and machine learning to help identify fake news (Okoro et al. 2018). Humans only have a 4% chance of identifying fake news if they take a guess and can only identify fake news 54% of the time (Okoro et al. 2018). The hybrid model as proven to increase this percentage (Okoro et al. 2018). To make the hybrid model effective it combines social media news with machine learning and a network approach (Okoro et al. 2018). The purpose of this model is to identify the probability that the news could be fake (Okoro et al. 2018). Another hybrid model called CSI (capture, score, integrate) has been developed and functions on the main elements; (1) capture - the process of extracting representations of articles by using a Recurrent Neutral Network (RNN), (2) Score – to create a score and representation vector, (3) Integrate – to integrate the outputs of the capture and score resulting in a vector which is used for classification (Ruchansky et al. 2017).

## 5 Conclusion

In this paper we discussed the prevalence of fake news and how technology has changed over the last years enabling us to develop tools that can be used in the fight against fake news. We also explored the importance of identifying fake news, the influence that misinformation can have on the public's decision making and which approaches exist to combat fake news. The current battle against fake news on COVID-19 and the uncertainty surrounding it, shows that a hybrid approach towards fake news detection is needed. Human wisdom as well as digital tools need to be harnessed in this process. Hopefully some of these measures will stay in place and that digital media platform owners and public will take responsibility and work together in detecting and combatting fake news.

## References

- Ahmed, S., Hinkelmann, K., Corradini, F.: Combining machine learning with knowledge engineering to detect fake news in social networks - a survey. In: Proceedings of the AAAI 2019 Spring Symposium, vol. 12 (2019)
- Albright, J.: Welcome to the era of fake news. *Media Commun.* **5**(2), 87 (2017). <https://doi.org/10.17645/mac.v5i2.977>
- Atodiresei, C.-S., Tănăselea, A., Iftene, A.: Identifying fake news and fake users on Twitter. *Procedia Comput. Sci.* **126**, 451–461 (2018). <https://doi.org/10.1016/j.procs.2018.07.279>
- Burkhardt, J.M.: History of fake news. *Libr. Technol. Rep.* **53**(8), 37 (2017)
- Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., Freire, J.: A topic-agnostic approach for identifying fake news pages. In: Companion Proceedings of the 2019 World Wide Web Conference on - WWW 2019, pp. 975–980 (2019). <https://doi.org/10.1145/3308560.3316739>

- Chen, Y., Conroy, N.J., Rubin, V.L.: Misleading online content: recognizing clickbait as false news? In: Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection - WMDD 2015, Seattle, Washington, USA, pp. 15–19. ACM Press (2015a). <https://doi.org/10.1145/2823465.2823467>
- Chen, Y., Conroy, N.J., Rubin, V.L.: News in an online world: the need for an ‘automatic crap detector.’ Proc. Assoc. Inf. Sci. Technol. **52**(1), 1–4 (2015b). <https://doi.org/10.1002/pr2.2015.145052010081>
- 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 - KDD 2017, Halifax, NS, Canada, pp. 1803–1812. ACM Press (2017). <https://doi.org/10.1145/3097983.3098131>
- 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: International AAAI Conference on Web and Social Media, vol. 8 (2017)
- Macaulay, T.: Can technology solve the fake news problem it helped create? (2018). <https://www.techworld.com/startups/can-technology-solve-fake-news-problem-it-helped-create-3672139/>
- Marr, B.: Coronavirus fake news: how Facebook, Twitter, and Instagram are tackling the problem. Forbes (2020). <https://www.forbes.com/sites/bernardmarr/2020/03/27/finding-the-truth-about-covid-19-how-facebook-twitter-and-instagram-are-tackling-fake-news/>
- Okoro, E.M., Abara, B.A., Umagba, A.O., Ajonye, A.A., Isa, Z.S.: A hybrid approach to fake news detection on social media. Niger. J. Technol. **37**(2), 454 (2018). <https://doi.org/10.4314/njt.v37i2.22>
- Paskin, D.: Real or fake news: who knows? J. Soc. Media Soc. **7**(2), 252–273 (2018)
- Pennycook, G., Rand, D.G.: Fighting misinformation on social media using crowdsourced judgments of news source quality. Proc. Natl. Acad. Sci. **116**(7), 2521–2526 (2019). <https://doi.org/10.1073/pnas.1806781116>
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. In: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, pp. 3391–3401. Association for Computational Linguistics (2018). <https://www.aclweb.org/anthology/C18-1287>
- Pothast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B.: A stylometric inquiry into hyperpartisan and fake news (2017). arXiv Preprint [arXiv:1702.05638](https://arxiv.org/abs/1702.05638)
- Qazvinian, V., Rosengren, E., Radev, D.R., Mei, Q.: Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 1589–1599 (2011)
- Ruchansky, N., Seo, S., Liu, Y.: CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, pp. 797–806 (2017). <https://doi.org/10.1145/3132847.3132877>
- Schade, U.: Software that can automatically detect fake news. Comput. Sci. Eng. **3** (2019)
- Shao, C., Ciampaglia, G.L., Varol, O., Yang, K., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots. Nat. Commun. **9**(1), 4787 (2018). <https://doi.org/10.1038/s41467-018-06930-7>
- Sivasangari, V., Anand, P.V., Santhya, R.: A modern approach to identify the fake news using machine learning. Int. J. Pure Appl. Math. **118**(20), 10 (2018)
- Sparks, H., Frishberg, H.: Facebook gives step-by-step instructions on how to spot fake news (2020). <https://nypost.com/2020/03/26/facebook-gives-step-by-step-instructions-on-how-to-spot-fake-news/>
- Stahl, K.: Fake news detection in social media. California State University Stanislaus, 6 (2018)
- Stein-Smith, K.: Librarians, information literacy, and fake news. Strat. Libr. **37** (2017)

- Thota, A., Tilak, P., Ahluwalia, S., Lohia, N.: Fake news detection: a deep learning approach. *SMU Data Sci. Rev.* **1**(3), 21 (2018)
- 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, pp. 18–22. Association for Computational Linguistics (2014). <https://doi.org/10.3115/v1/W14-2508>
- Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., Yu, P.S.: TI-CNN: convolutional neural networks for fake news detection (2018). arXiv preprint [arXiv:1806.00749](https://arxiv.org/abs/1806.00749)
- Zhou, X., Zafarani, R.: Fake news: a survey of research, detection methods, and opportunities (2018). arXiv preprint [arXiv:1812.00315](https://arxiv.org/abs/1812.00315)