Social Network Analysis and Mining (2023) 13:30
https://doi.org/10.1007/s13278-023-01028-5

ORIGINAL ARTICLE

Fake news, disinformation and misinformation in social media: a review

Esma Aïmeur[1] · **Sabrine Amri**[1] · **Gilles Brassard**[1]

Received: 20 October 2022 / Revised: 7 January 2023 / Accepted: 12 January 2023 / Published online: 9 February 2023

Abstract
Online social networks (OSNs) are rapidly growing and have become a huge source of all kinds of global and local news for millions of users. However, OSNs are a double-edged sword. Although the great advantages they offer such as unlimited easy communication and instant news and information, they can also have many disadvantages and issues. One of their major challenging issues is the spread of fake news. Fake news identification is still a complex unresolved issue. Furthermore, fake news detection on OSNs presents unique characteristics and challenges that make finding a solution anything but trivial. On the other hand, artificial intelligence (AI) approaches are still incapable of overcoming this challenging problem. To make matters worse, AI techniques such as machine learning and deep learning are leveraged to deceive people by creating and disseminating fake content. Consequently, automatic fake news detection remains a huge challenge, primarily because the content is designed in a way to closely resemble the truth, and it is often hard to determine its veracity by AI alone without additional information from third parties. This work aims to provide a comprehensive and systematic review of fake news research as well as a fundamental review of existing approaches used to detect and prevent fake news from spreading via OSNs. We present the research problem and the existing challenges, discuss the state of the art in existing approaches for fake news detection, and point out the future research directions in tackling the challenges.

Keywords Fake news · Disinformation · Misinformation · Information disorder · Online deception · Online social networks

1 Introduction

1.1 Context and motivation

Fake news, disinformation and misinformation have become such a scourge that Marcia McNutt, president of the National Academy of Sciences of the United States, is quoted to have said (making an implicit reference to the COVID-19 pandemic) "Misinformation is worse than an epidemic: It spreads at the speed of light throughout the globe and can prove deadly when it reinforces misplaced personal bias against all trustworthy evidence" in a joint statement of the National Academies[1] posted on July 15, 2021. Indeed, although online social networks (OSNs), also called social media, have improved the ease with which real-time information is broadcast; its popularity and its massive use have

✉ Sabrine Amri
sabrine.amri@umontreal.ca

Esma Aïmeur
aimeur@iro.umontreal.ca

Gilles Brassard
brassard@iro.umontreal.ca

[1] Department of Computer Science and Operations Research (DIRO), University of Montreal, Montreal, Canada

[1] https://www.nationalacademies.org/news/2021/07/as-surgeon-general-urges-whole-of-society-effort-to-fight-health-misinformation-the-work-of-the-national-academies-helps-foster-an-evidence-based-information-environment, last access date: 26-12-2022.

expanded the spread of fake news by increasing the speed and scope at which it can spread. Fake news may refer to the manipulation of information that can be carried out through the production of false information, or the distortion of true information. However, that does not mean that this problem is only created with social media. A long time ago, there were rumors in the traditional media that Elvis was not dead,[2] that the Earth was flat,[3] that aliens had invaded us,[4], etc.

Therefore, social media has become nowadays a powerful source for fake news dissemination (Sharma et al. 2019; Shu et al. 2017). According to Pew Research Center's analysis of the news use across social media platforms, in 2020, about half of American adults get news on social media at least sometimes,[5] while in 2018, only one-fifth of them say they often get news via social media.[6]

Hence, fake news can have a significant impact on society as manipulated and false content is easier to generate and harder to detect (Kumar and Shah 2018) and as disinformation actors change their tactics (Kumar and Shah 2018; Micallef et al. 2020). In 2017, Snow predicted in the MIT *Technology Review* (Snow 2017) that most individuals in mature economies will consume more false than valid information by 2022.

Recent news on the COVID-19 pandemic, which has flooded the web and created panic in many countries, has been reported as fake.[7] For example, holding your breath for ten seconds to one minute is not a self-test for COVID-19[8] (see Fig. 1). Similarly, online posts claiming to reveal various "cures" for COVID-19 such as eating boiled garlic or drinking chlorine dioxide (which is an industrial bleach), were verified[9] as fake and in some cases as dangerous and will never cure the infection.

Social media outperformed television as the major news source for young people of the UK and the USA.[10] Moreover, as it is easier to generate and disseminate news online than with traditional media or face to face, large volumes of fake news are produced online for many reasons (Shu et al. 2017). Furthermore, it has been reported in a previous study about the spread of online news on Twitter (Vosoughi et al. 2018) that the spread of false news online is six times faster than truthful content and that 70% of the users could not distinguish real from fake news (Vosoughi et al. 2018) due to the attraction of the novelty of the latter (Bovet and Makse 2019). It was determined that falsehood spreads significantly farther, faster, deeper and more broadly than the truth in all categories of information, and the effects are more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information (Vosoughi et al. 2018).

Over 1 million tweets were estimated to be related to fake news by the end of the 2016 US presidential election.[11] In 2017, in Germany, a government spokesman affirmed: "We are dealing with a phenomenon of a dimension that we have not seen before," referring to an unprecedented spread of fake news on social networks.[12] Given the strength of this new phenomenon, fake news has been chosen as the word of the year by the Macquarie dictionary both in 2016[13] and in 2018[14] as well as by the Collins dictionary in 2017.[15,16] Since 2020, the new term "infodemic" was coined, reflecting widespread researchers' concern (Gupta et al. 2022; Apuke and Omar 2021; Sharma et al. 2020; Hartley and Vu 2020; Micallef et al. 2020) about the proliferation of misinformation linked to the COVID-19 pandemic.

The Gartner Group's top strategic predictions for 2018 and beyond included the need for IT leaders to quickly develop Artificial Intelligence (AI) algorithms to address counterfeit reality and fake news.[17] However, fake news identification is a complex issue. (Snow 2017) questioned

[2] https://time.com/4897819/elvis-presley-alive-conspiracy-theories/, last access date: 26-12-2022.

[3] https://www.therichest.com/shocking/the-evidence-15-reasons-people-think-the-earth-is-flat/, last access date: 26-12-2022.

[4] https://www.grunge.com/657584/the-truth-about-1952s-alien-invasion-of-washington-dc/, last access date: 26-12-2022.

[5] https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/, last access date: 26-12-2022.

[6] https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/, last access date: 26-12-2022.

[7] https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes, last access date: 26-12-2022.

[8] https://www.factcheck.org/2020/03/viral-social-media-posts-offer-false-coronavirus-tips/, last access date: 26-12-2022.

[9] https://www.factcheck.org/2020/02/fake-coronavirus-cures-part-2-garlic-isnt-a-cure/, last access date: 26-12-2022.

[10] https://www.bbc.com/news/uk-36528256, last access date: 26-12-2022.

[11] https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory, last access date: 26-12-2022.

[12] https://www.theguardian.com/world/2017/jan/09/germany-investigating-spread-fake-news-online-russia-election, last access date: 26-12-2022.

[13] https://www.macquariedictionary.com.au/resources/view/word/of/the/year/2016, last access date: 26-12-2022.

[14] https://www.macquariedictionary.com.au/resources/view/word/of/the/year/2018, last access date: 26-12-2022.

[15] https://apnews.com/article/47466c5e260149b1a23641b9e319fda6, last access date: 26-12-2022.

[16] https://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist/, last access date: 26-12-2022.

[17] https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/, last access date: 26-12-2022.

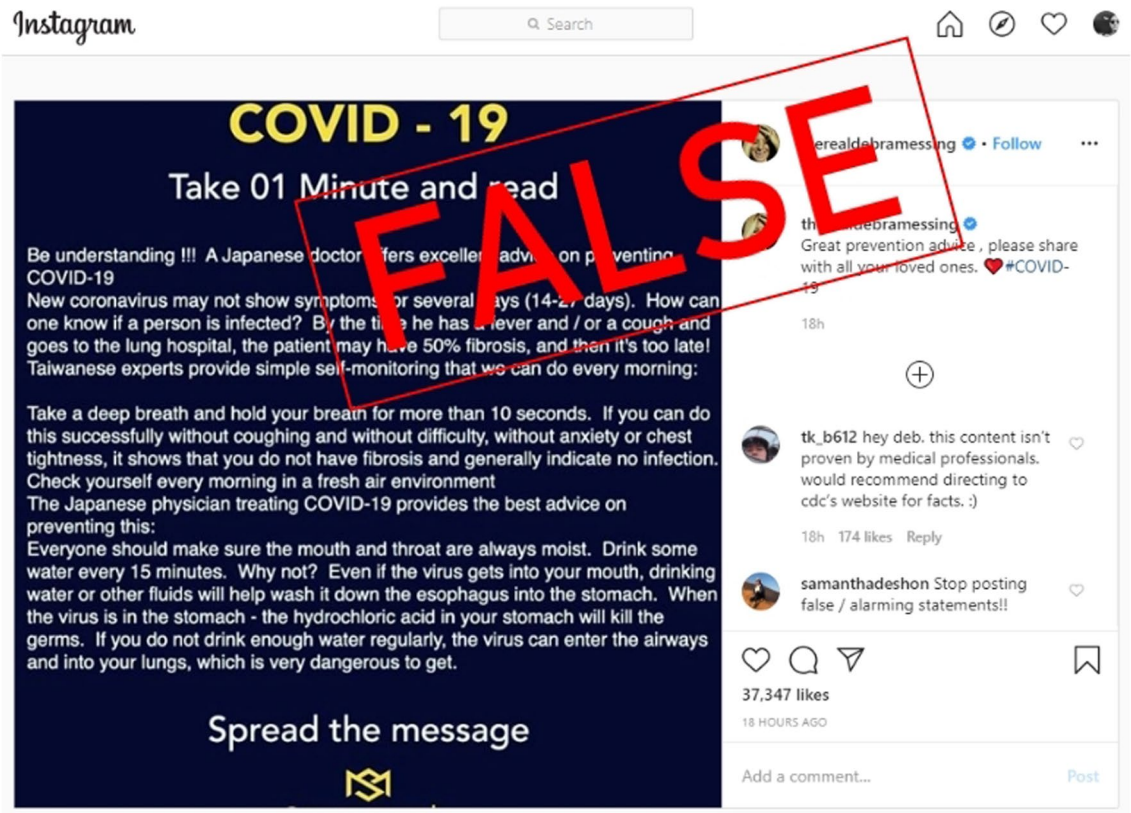

Fig. 1 Fake news example about a self-test for COVID-19 source: https://cdn.factcheck.org/UploadedFiles/Screenshot031120_false.jpg, last access date: 26-12-2022

the ability of AI to win the war against fake news. Similarly, other researchers concurred that even the best AI for spotting fake news is still ineffective.[18] Besides, recent studies have shown that the power of AI algorithms for identifying fake news is lower than its ability to create it Paschen (2019). Consequently, automatic fake news detection remains a huge challenge, primarily because the content is designed to closely resemble the truth in order to deceive users, and as a result, it is often hard to determine its veracity by AI alone. Therefore, it is crucial to consider more effective approaches to solve the problem of fake news in social media.

1.2 Contribution

The fake news problem has been addressed by researchers from various perspectives related to different topics. These topics include, but are not restricted to, *social science studies*, which investigate why and who falls for fake news (Altay et al. 2022; Batailler et al. 2022; Sterret et al. 2018; Badawy et al. 2019; Pennycook and Rand 2020; Weiss et al. 2020; Guadagno and Guttieri 2021), whom to trust and how perceptions of misinformation and disinformation relate to media trust and media consumption patterns (Hameleers et al. 2022), how fake news differs from personal lies (Chiu and Oh 2021; Escolà-Gascón 2021), examine how can the law regulate digital disinformation and how governments can regulate the values of social media companies that themselves regulate disinformation spread on their platforms (Marsden et al. 2020; Schuyler 2019; Vasu et al. 2018; Burshtein 2017; Waldman 2017; Alemanno 2018; Verstraete et al. 2017), and argue the challenges to democracy (Jungherr and Schroeder 2021); *Behavioral interventions studies*, which examine what literacy ideas mean in the age of dis/mis- and malinformation (Carmi et al. 2020), investigate whether media literacy helps identification of fake news (Jones-Jang et al. 2021) and attempt to improve people's news literacy (Apuke et al. 2022; Dame Adjin-Tettey 2022; Hameleers 2022; Nagel 2022; Jones-Jang et al. 2021; Mihailidis and Viotty 2017; García et al. 2020) by encouraging people to pause to assess credibility of headlines (Fazio 2020), promote civic online reasoning (McGrew 2020; McGrew et al. 2018) and critical thinking (Lutzke et al. 2019), together with evaluations of credibility indicators (Bhuiyan et al. 2020; Nygren et al. 2019; Shao et al. 2018a; Pennycook et al. 2020a, b; Clayton et al. 2020; Ozturk et al.

[18] https://www.technologyreview.com/s/612236/even-the-best-ai-for-spotting-fake-news-is-still-terrible/, last access date: 26-12-2022.

2015; Metzger et al. 2020; Sherman et al. 2020; Nekmat 2020; Brashier et al. 2021; Chung and Kim 2021; Lanius et al. 2021); as well as *social media-driven studies*, which investigate the effect of signals (e.g., sources) to detect and recognize fake news (Vraga and Bode 2017; Jakesch et al. 2019; Shen et al. 2019; Avram et al. 2020; Hameleers et al. 2020; Dias et al. 2020; Nyhan et al. 2020; Bode and Vraga 2015; Tsang 2020; Vishwakarma et al. 2019; Yavary et al. 2020) and investigate fake and reliable news sources using complex networks analysis based on search engine optimization metric (Mazzeo and Rapisarda 2022).

The impacts of fake news have reached various areas and disciplines beyond online social networks and society (García et al. 2020) such as economics (Clarke et al. 2020; Kogan et al. 2019; Goldstein and Yang 2019), psychology (Roozenbeek et al. 2020a; Van der Linden and Roozenbeek 2020; Roozenbeek and van der Linden 2019), political science (Valenzuela et al. 2022; Bringula et al. 2022; Ricard and Medeiros 2020; Van der Linden et al. 2020; Allcott and Gentzkow 2017; Grinberg et al. 2019; Guess et al. 2019; Baptista and Gradim 2020), health science (Alonso-Galbán and Alemañy-Castilla 2022; Desai et al. 2022; Apuke and Omar 2021; Escolà-Gascón 2021; Wang et al. 2019c; Hartley and Vu 2020; Micallef et al. 2020; Pennycook et al. 2020b; Sharma et al. 2020; Roozenbeek et al. 2020b), environmental science (e.g., climate change) (Treen et al. 2020; Lutzke et al. 2019; Lewandowsky 2020; Maertens et al. 2020), etc.

Interesting research has been carried out to review and study the fake news issue in online social networks. Some focus not only on fake news, but also distinguish between fake news and rumor (Bondielli and Marcelloni 2019; Meel and Vishwakarma 2020), while others tackle the whole problem, from characterization to processing techniques (Shu et al. 2017; Guo et al. 2020; Zhou and Zafarani 2020). However, they mostly focus on studying approaches from a machine learning perspective (Bondielli and Marcelloni 2019), data mining perspective (Shu et al. 2017), crowd intelligence perspective (Guo et al. 2020), or knowledge-based perspective (Zhou and Zafarani 2020). Furthermore, most of these studies ignore at least one of the mentioned perspectives, and in many cases, they do not cover other existing detection approaches using methods such as blockchain and fact-checking, as well as analysis on metrics used for Search Engine Optimization (Mazzeo and Rapisarda 2022). However, in our work and to the best of our knowledge, we cover all the approaches used for fake news detection. Indeed, we investigate the proposed solutions from broader perspectives (i.e., the detection techniques that are used, as well as the different aspects and types of the information used).

Therefore, in this paper, we are highly motivated by the following facts. First, fake news detection on social media is still in the early age of development, and many challenging issues remain that require deeper investigation. Hence, it is necessary to discuss potential research directions that can improve fake news detection and mitigation tasks. However, the dynamic nature of fake news propagation through social networks further complicates matters (Sharma et al. 2019). False information can easily reach and impact a large number of users in a short time (Friggeri et al. 2014; Qian et al. 2018). Moreover, fact-checking organizations cannot keep up with the dynamics of propagation as they require human verification, which can hold back a timely and cost-effective response (Kim et al. 2018; Ruchansky et al. 2017; Shu et al. 2018a).

Our work focuses primarily on understanding the "fake news" problem, its related challenges and root causes, and reviewing automatic fake news detection and mitigation methods in online social networks as addressed by researchers. The main contributions that differentiate us from other works are summarized below:

- We present the general context from which the fake news problem emerged (i.e., online deception)
- We review existing definitions of fake news, identify the terms and features most commonly used to define fake news, and categorize related works accordingly.
- We propose a fake news typology classification based on the various categorizations of fake news reported in the literature.
- We point out the most challenging factors preventing researchers from proposing highly effective solutions for automatic fake news detection in social media.
- We highlight and classify representative studies in the domain of automatic fake news detection and mitigation on online social networks including the key methods and techniques used to generate detection models.
- We discuss the key shortcomings that may inhibit the effectiveness of the proposed fake news detection methods in online social networks.
- We provide recommendations that can help address these shortcomings and improve the quality of research in this domain.

The rest of this article is organized as follows. We explain the methodology with which the studied references are collected and selected in Sect. 2. We introduce the online deception problem in Sect. 3. We highlight the modern-day problem of fake news in Sect. 4, followed by challenges facing fake news detection and mitigation tasks in Sect. 5. We provide a comprehensive literature review of the most relevant scholarly works on fake news detection in Sect. 6. We provide a critical discussion and recommendations that may fill some of the gaps we have identified, as well as a classification of the reviewed automatic fake news detection

approaches, in Sect. 7. Finally, we provide a conclusion and propose some future directions in Sect. 8.

2 Review methodology

This section introduces the systematic review methodology on which we relied to perform our study. We start with the formulation of the research questions, which allowed us to select the relevant research literature. Then, we provide the different sources of information together with the search and inclusion/exclusion criteria we used to select the final set of papers.

2.1 Research questions formulation

The research scope, research questions, and inclusion/exclusion criteria were established following an initial evaluation of the literature and the following research questions were formulated and addressed.

- RQ1: what is fake news in social media, how is it defined in the literature, what are its related concepts, and the different types of it?
- RQ2: What are the existing challenges and issues related to fake news?
- RQ3: What are the available techniques used to perform fake news detection in social media?

2.2 Sources of information

We broadly searched for journal and conference research articles, books, and magazines as a source of data to extract relevant articles. We used the main sources of scientific databases and digital libraries in our search, such as Google Scholar,[19] IEEE Xplore,[20] Springer Link,[21] ScienceDirect,[22] Scopus,[23] ACM Digital Library.[24] Also, we screened most of the related high-profile conferences such as WWW, SIGKDD, VLDB, ICDE and so on to find out the recent work.

2.3 Search criteria

We focused our research over a period of ten years, but we made sure that about two-thirds of the research papers that we considered were published in or after 2019. Additionally, we defined a set of keywords to search the above-mentioned scientific databases since we concentrated on reviewing the current state of the art in addition to the challenges and the future direction. The set of keywords includes the following terms: fake news, disinformation, misinformation, information disorder, social media, detection techniques, detection methods, survey, literature review.

Table 1 List of keywords for searching relevant articles

Fake news + social media
Fake news + disinformation
Fake news + misinformation
Fake news + information disorder
Fake news + survey
Fake news + detection methods
Fake news + literature review
Fake news + detection techniques
Fake news + detection + social media
Disinformation + misinformation + social media

2.4 Study selection, exclusion and inclusion criteria

To retrieve relevant research articles, based on our sources of information and search criteria, a systematic keyword-based search was carried out by posing different search queries, as shown in Table 1.

We discovered a primary list of articles. On the obtained initial list of studies, we applied a set of inclusion/exclusion criteria presented in Table 2 to select the appropriate research papers. The inclusion and exclusion principles are applied to determine whether a study should be included or not.

After reading the abstract, we excluded some articles that did not meet our criteria. We chose the most important research to help us understand the field. We reviewed the articles completely and found only 61 research papers that discuss the definition of the term fake news and its related concepts (see Table 4). We used the remaining papers to understand the field, reveal the challenges, review the detection techniques, and discuss future directions.

3 A brief introduction of online deception

The Cambridge Online Dictionary defines Deception as "*the act of hiding the truth, especially to get an advantage*." Deception relies on peoples' trust, doubt and strong emotions that may prevent them from thinking and acting clearly (Aïmeur et al. 2018). We also define it in previous work (Aïmeur et al. 2018) as the process that undermines the ability to consciously make decisions and take convenient actions, following personal values and boundaries. In

[19] https://scholar.google.ca/, last access date: 26-12-2022.
[20] https://ieeexplore.ieee.org/, last access date: 26-12-2022.
[21] https://link.springer.com/, last access date: 26-12-2022.
[22] https://www.sciencedirect.com/, last access date: 26-12-2022.
[23] https://www.scopus.com/, last access date: 26-12-2022.
[24] https://www.acm.org/digital-library, last access date: 26-12-2022.

Table 2 Inclusion and exclusion criteria

Inclusion criterion	Exclusion criterion
Peer-reviewed and written in the English language	Articles in a different language than English.
Clearly describes fake news, misinformation and disinformation problems in social networks	Does not focus on fake news, misinformation, or disinformation problem in social networks
Written by academic or industrial researchers	Short papers, posters or similar
Have a high number of citations	
Recent articles only (last ten years)	
In the case of equivalent studies, the one published in the highest-rated journal or conference is selected to sustain a high-quality set of articles on which the review is conducted	Articles not following these inclusion criteria
Articles that propose methodologies, methods, or approaches for fake news detection online social networks	

other words, deception gets people to do things they would not otherwise do. In the context of online deception, several factors need to be considered: the deceiver, the purpose or aim of the deception, the social media service, the deception technique and the potential target (Aïmeur et al. 2018; Hage et al. 2021).

Researchers are working on developing new ways to protect users and prevent online deception (Aïmeur et al. 2018). Due to the sophistication of attacks, this is a complex task. Hence, malicious attackers are using more complex tools and strategies to deceive users. Furthermore, the way information is organized and exchanged in social media may lead to exposing OSN users to many risks (Aïmeur et al. 2013).

In fact, this field is one of the recent research areas that need collaborative efforts of multidisciplinary practices such as psychology, sociology, journalism, computer science as well as cyber-security and digital marketing (which are not yet well explored in the field of dis/mis/malinformation but relevant for future research). Moreover, Ismailov et al. (2020) analyzed the main causes that could be responsible for the efficiency gap between laboratory results and real-world implementations.

In this paper, it is not in our scope of work to review online deception state of the art. However, we think it is crucial to note that fake news, misinformation and disinformation are indeed parts of the larger landscape of online deception (Hage et al. 2021).

4 Fake news, the modern-day problem

Fake news has existed for a very long time, much before their wide circulation became facilitated by the invention of the printing press.[25] For instance, Socrates was condemned to death more than twenty-five hundred years ago under the fake news that he was guilty of impiety against the pantheon of Athens and corruption of the youth.[26] A Google Trends Analysis of the term "fake news" reveals an explosion in popularity around the time of the 2016 US presidential election.[27] Fake news detection is a problem that has recently been addressed by numerous organizations, including the European Union[28] and NATO.[29]

In this section, we first overview the fake news definitions as they were provided in the literature. We identify the terms and features used in the definitions, and we classify the latter based on them. Then, we provide a fake news typology based on distinct categorizations that we propose, and we define and compare the most cited forms of one specific fake news category (i.e., the intent-based fake news category).

4.1 Definitions of fake news

"Fake news" is defined in the Collins English Dictionary as false and often sensational information disseminated under the guise of news reporting,[30] yet the term has evolved over time and has become synonymous with the spread of false information (Cooke 2017).

The first definition of the term *fake news* was provided by Allcott and Gentzkow (2017) as news articles that are intentionally and verifiably false and could mislead readers. Then, other definitions were provided in the literature, but they all agree on the *authenticity* of fake news to be false (i.e., being

[25] https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535, last access date: 26-12-2022.

[26] https://en.wikipedia.org/wiki/Trial_of_Socrates, last access date: 26-12-2022.

[27] https://trends.google.com/trends/explore?hl=en-US &tz=-180 &date=2013-12-06+2018-01-06 &geo=US &q=fake+news &sni=3, last access date: 26-12-2022.

[28] https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation, last access date: 26-12-2022.

[29] https://www.nato.int/cps/en/natohq/177273.htm, last access date: 26-12-2022.

[30] https://www.collinsdictionary.com/dictionary/english/fake-news, last access date: 26-12-2022.

non-factual). However, they disagree on the inclusion and exclusion of some related concepts such as *satire*, *rumors*, *conspiracy theories*, *misinformation* and *hoaxes* from the given definition. More recently, Nakov (2020) reported that the term fake news started to mean different things to different people, and for some politicians, it even means "news that I do not like."

Hence, there is still no agreed definition of the term "fake news." Moreover, we can find many terms and concepts in the literature that refer to fake news (Van der Linden et al. 2020; Molina et al. 2021) (Abu Arqoub et al. 2022; Allen et al. 2020; Allcott and Gentzkow 2017; Shu et al. 2017; Sharma et al. 2019; Zhou and Zafarani 2020; Zhang and Ghorbani 2020; Conroy et al. 2015; Celliers and Hattingh 2020; Nakov 2020; Shu et al. 2020c; Jin et al. 2016; Rubin et al. 2016; Balmas 2014; Brewer et al. 2013; Egelhofer and Lecheler 2019; Mustafaraj and Metaxas 2017; Klein and Wueller 2017; Potthast et al. 2017; Lazer et al. 2018; Weiss et al. 2020; Tandoc Jr et al. 2021; Guadagno and Guttieri 2021), disinformation (Kapantai et al. 2021; Shu et al. 2020a, c; Kumar et al. 2016; Bhattacharjee et al. 2020; Marsden et al. 2020; Jungherr and Schroeder 2021; Starbird et al. 2019; Ireton and Posetti 2018), misinformation (Wu et al. 2019; Shu et al. 2020c; Shao et al. 2016, 2018b; Pennycook and Rand 2019; Micallef et al. 2020), malinformation (Dame Adjin-Tettey 2022) (Carmi et al. 2020; Shu et al. 2020c), false information (Kumar and Shah 2018; Guo et al. 2020; Habib et al. 2019), information disorder (Shu et al. 2020c; Wardle and Derakhshan 2017; Wardle 2018; Derakhshan and Wardle 2017), information warfare (Guadagno and Guttieri 2021) and information pollution (Meel and Vishwakarma 2020).

There is also a remarkable amount of disagreement over the classification of the term fake news in the research literature, as well as in policy (de Cock Buning 2018; ERGA 2018, 2021). Some consider fake news as a type of misinformation (Allen et al. 2020; Singh et al. 2021; Ha et al. 2021; Pennycook and Rand 2019; Shao et al. 2018b; Di Domenico et al. 2021; Sharma et al. 2019; Celliers and Hattingh 2020; Klein and Wueller 2017; Potthast et al. 2017; Islam et al. 2020), others consider it as a type of disinformation (de Cock Buning 2018) (Bringula et al. 2022; Baptista and Gradim 2022; Tsang 2020; Tandoc Jr et al. 2021; Bastick 2021; Khan et al. 2019; Shu et al. 2017; Nakov 2020; Shu et al. 2020c; Egelhofer and Lecheler 2019), while others associate the term with both disinformation and misinformation (Wu et al. 2022; Dame Adjin-Tettey 2022; Hameleers et al. 2022; Carmi et al. 2020; Allcott and Gentzkow 2017; Zhang and Ghorbani 2020; Potthast et al. 2017; Weiss et al. 2020; Tandoc Jr et al. 2021; Guadagno and Guttieri 2021). On the other hand, some prefer to differentiate fake news from both terms (ERGA 2018; Molina et al. 2021; ERGA 2021) (Zhou and Zafarani 2020; Jin et al. 2016; Rubin et al. 2016; Balmas 2014; Brewer et al. 2013).

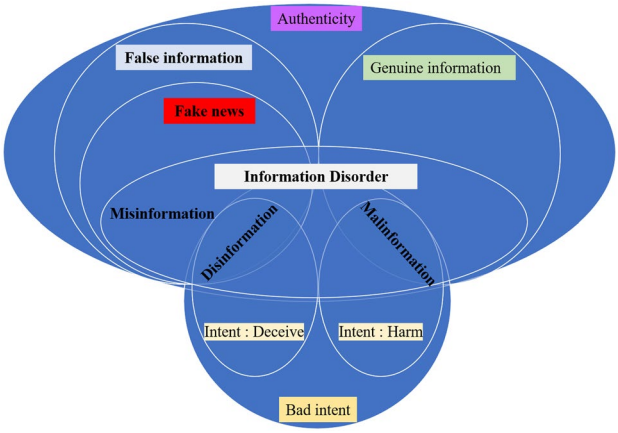

Fig. 2 Modeling of the relationship between terms related to fake news

The existing terms can be separated into two groups. The first group represents the general terms, which are *information disorder*, *false information* and *fake news*, each of which includes a subset of terms from the second group. The second group represents the elementary terms, which are *misinformation*, *disinformation* and *malinformation*. The literature agrees on the definitions of the latter group, but there is still no agreed-upon definition of the first group. In Fig. 2, we model the relationship between the most used terms in the literature.

The terms most used in the literature to refer, categorize and classify fake news can be summarized and defined as shown in Table 3, in which we capture the similarities and show the differences between the different terms based on two common key features, which are the intent and the authenticity of the news content. The intent feature refers to the intention behind the term that is used (i.e., whether or not the purpose is to mislead or cause harm), whereas the authenticity feature refers to its factual aspect. (i.e., whether the content is verifiably false or not, which we label as genuine in the second case). Some of these terms are explicitly used to refer to fake news (i.e., disinformation, misinformation and false information), while others are not (i.e., malinformation). In the comparison table, the empty dash (–) cell denotes that the classification does not apply.

In Fig. 3, we identify the different features used in the literature to define fake news (i.e., intent, authenticity and knowledge). Hence, some definitions are based on two key features, which are *authenticity and intent* (i.e., news articles that are intentionally and verifiably false and could mislead readers). However, other definitions are based on either authenticity *or* intent. Other researchers categorize

Table 3 A comparison between used terms based on intent and authenticity

Term	Definition	Intent	Authenticity
False information	Verifiably false information	–	False
Misinformation	False information that is shared without the intention to mislead or to cause harm	Not to mislead	False
Disinformation	False information that is shared to intentionally mislead	To mislead	False
Malinformation	Genuine information that is shared with an intent to cause harm	To cause harm	Genuine

false information on the web and social media based on its intent and *knowledge* (i.e., when there is a single ground truth). In Table 4, we classify the existing fake news definitions based on the used *term* and the used *features*. In the classification, the references in the cells refer to the research study in which a fake news definition was provided, while the empty dash (–) cells denote that the classification does not apply.

4.2 Fake news typology

Various categorizations of fake news have been provided in the literature. We can distinguish two major categories of fake news based on the studied perspective (i.e., intention or content) as shown in Fig. 4. However, our proposed fake news typology is not about detection methods, and it is not exclusive. Hence, a given category of fake news can be described based on both perspectives (i.e., intention and content) at the same time. For instance, satire (i.e., intent-based fake news) can contain text and/or multimedia content types of data (e.g., headline, body, image, video) (i.e., content-based fake news) and so on.

Most researchers classify fake news based on the intent (Collins et al. 2020; Bondielli and Marcelloni 2019; Zannettou et al. 2019; Kumar et al. 2016; Wardle 2017; Shu et al. 2017; Kumar and Shah 2018) (see Sect. 4.2.2). However, other researchers (Parikh and Atrey 2018; Fraga-Lamas and Fernández-Caramés 2020; Hasan and Salah 2019; Masciari et al. 2020; Bakdash et al. 2018; Elhadad et al. 2019; Yang et al. 2019b) focus on the content to categorize types of fake news through distinguishing the different formats and content types of data in the news (e.g., text and/or multimedia).

Recently, another classification was proposed by Zhang and Ghorbani (2020). It is based on the combination of content and intent to categorize fake news. They distinguish physical news content and non-physical news content from fake news. Physical content consists of the carriers and format of the news, and non-physical content consists of the opinions, emotions, attitudes and sentiments that the news creators want to express.

4.2.1 Content-based fake news category

According to researchers of this category (Parikh and Atrey 2018; Fraga-Lamas and Fernández-Caramés 2020; Hasan and Salah 2019; Masciari et al. 2020; Bakdash et al. 2018; Elhadad et al. 2019; Yang et al. 2019b), forms of fake news may include false text such as hyperlinks or embedded content; multimedia such as false videos (Demuyakor and Opata 2022), images (Masciari et al. 2020; Shen et al. 2019), audios (Demuyakor and Opata 2022) and so on. Moreover, we can also find multimodal content (Shu et al. 2020a) that is fake news articles and posts composed of multiple types of data combined together, for example, a fabricated image along with a text related to the image (Shu et al. 2020a). In this category of fake news forms, we can mention as examples deepfake videos (Yang et al. 2019b) and GAN-generated fake images (Zhang et al. 2019b), which are artificial intelligence-based machine-generated fake content that are hard for unsophisticated social network users to identify.

The effects of these forms of fake news content vary on the credibility assessment, as well as sharing intentions which influences the spread of fake news on OSNs. For instance, people with little knowledge about the issue compared to those who are strongly concerned about the key issue of fake news tend to be easier to convince that the misleading or fake news is real, especially when shared via a video modality as compared to the text or the audio modality (Demuyakor and Opata 2022).

4.2.2 Intent-based Fake News Category

The most often mentioned and discussed forms of fake news according to researchers in this category include but are not restricted to *clickbait*, *hoax*, *rumor*, *satire*, *propaganda*, *framing*, *conspiracy theories* and others. In the following subsections, we explain these types of fake news as they were defined in the literature and undertake a brief comparison between them as depicted in Table 5. The following are the most cited forms of intent-based types of fake news, and their comparison is based on what we suspect are the most common criteria mentioned by researchers.

Table 4 Classification of fake news definitions based on the used term and features

	Fake news	Misinformation	Disinformation	False information	Malinformation	Information disorder
Intent and authenticity	Shu et al. (2017), Sharma et al. (2019), Mustafaraj and Metaxas (2017), Klein and Wueller (2017), Potthast et al. (2017), Allcott and Gentzkow (2017), Zhou and Zafarani (2020), Zhang and Ghorbani (2020), Conroy et al. (2015), Celliers and Hattingh (2020), Nakov (2020), Shu et al. (2020c), Tandoc Jr et al. (2021), Abu Arqoub et al. (2022),Molina et al. (2021), de Cock Buning (2018), Meel and Vishwakarma (2020)	Wu et al. (2019), Shu et al. (2020c), Islam et al. (2020), Hameleers et al. (2022)	Kapantai et al. (2021), Shu et al. (2020a), Shu et al. (2020c),Kumar et al. (2016), Jungherr and Schroeder (2021), Starbird et al. (2019), de Cock Buning (2018), Bastick (2021), Bringula et al. (2022), Tsang (2020), Hameleers et al. (2022), Wu et al. (2022)	–	Shu et al. (2020c), Di Domenico et al. (2021), Dame Adjin-Tettey (2022)	Wardle and Derakhshan (2017), Wardle Wardle (2018), Derakhshan and Wardle (2017), Shu et al. (2020c)
Intent or authenticity	Jin et al. (2016), Rubin et al. (2016), Balmas (2014),Brewer et al. (2013), Egelhofer and Lecheler (2019), Lazer et al. (2018), Allen et al. (2020), Guadagno and Guttieri (2021), Van der Linden et al. (2020), ERGA (2018)	Pennycook and Rand (2019), Shao et al. (2016), Shao et al. (2018b),Micallef et al. (2020), Ha et al. (2021), Singh et al. (2021), Wu et al. (2022)	Marsden et al. (2020), Ireton and Posetti (2018), ERGA (2021), Baptista and Gradim (2022)	Habib et al. (2019)	Carmi et al. (2020)	–
Intent and knowledge	Weiss et al. (2020)	–	Bhattacharjee et al. (2020), Khan et al. (2019)	Kumar and Shah (2018), Guo et al. (2020)	–	–

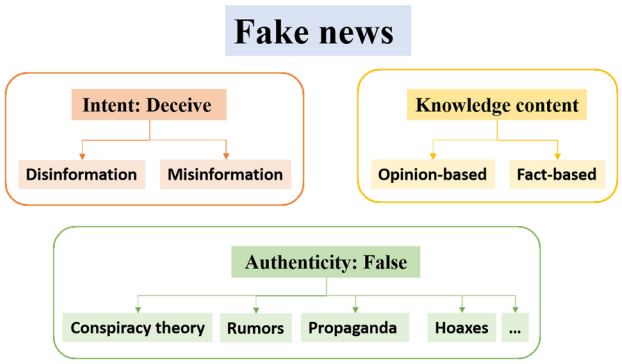

Fig. 3 The features used for fake news definition

Clickbait Clickbait refers to misleading headlines and thumbnails of content on the web (Zannettou et al. 2019) that tend to be fake stories with catchy headlines aimed at enticing the reader to click on a link (Collins et al. 2020). This type of fake news is considered to be the least severe type of false information because if a user reads/views the whole content, it is possible to distinguish if the headline and/or the thumbnail was misleading (Zannettou et al. 2019). However, the goal behind using clickbait is to increase the traffic to a website (Zannettou et al. 2019).

Hoax A hoax is a false (Zubiaga et al. 2018) or inaccurate (Zannettou et al. 2019) intentionally fabricated (Collins et al. 2020) news story used to masquerade the truth (Zubiaga et al. 2018) and is presented as factual (Zannettou et al. 2019) to deceive the public or audiences (Collins et al. 2020). This category is also known either as half-truth or factoid stories (Zannettou et al. 2019). Popular examples of hoaxes are stories that report the false death of celebrities (Zannettou et al. 2019) and public figures (Collins et al. 2020). Recently, hoaxes about the COVID-19 have been circulating through social media.

Rumor The term rumor refers to ambiguous or never confirmed claims (Zannettou et al. 2019) that are disseminated with a lack of evidence to support them (Sharma et al. 2019). This kind of information is widely propagated on OSNs (Zannettou et al. 2019). However, they are not necessarily false and may turn out to be true (Zubiaga et al. 2018). Rumors originate from unverified sources but may be true or false or remain unresolved (Zubiaga et al. 2018).

Satire Satire refers to stories that contain a lot of irony and humor (Zannettou et al. 2019). It presents stories as news that might be factually incorrect, but the intent is not to deceive but rather to call out, ridicule, or to expose behavior

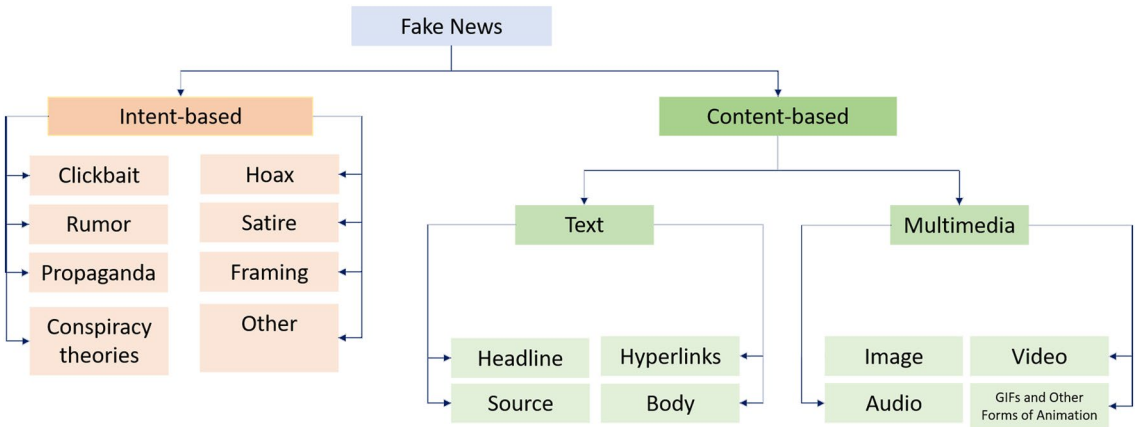

Fig. 4 Fake news typology

Table 5 A comparison between the different types of intent-based fake news

	Intent to deceive	Propagation	Negative Impact	Goal
Clickbait	High	Slow	Low	Popularity, Profit
Hoax	High	Fast	Low	Other
Rumor	High	Fast	High	Other
Satire	Low	Slow	Low	Popularity, Other
Propaganda	High	Fast	High	Popularity
Framing	High	Fast	Low	Other
Conspiracy theory	High	Fast	High	Other

that is shameful, corrupt, or otherwise "bad" (Golbeck et al. 2018). This is done with a fabricated story or by exaggerating the truth reported in mainstream media in the form of comedy (Collins et al. 2020). The intent behind satire seems kind of legitimate and many authors (such as Wardle (Wardle 2017)) do include satire as a type of fake news as there is no intention to cause harm but it has the potential to mislead or fool people.

Also, Golbeck et al. (2018) mention that there is a spectrum from fake to satirical news that they found to be exploited by many fake news sites. These sites used disclaimers at the bottom of their webpages to suggest they were "satirical" even when there was nothing satirical about their articles, to protect them from accusations about being fake. The difference with a satirical form of fake news is that the authors or the host present themselves as a comedian or as an entertainer rather than a journalist informing the public (Collins et al. 2020). However, most audiences believed the information passed in this satirical form because the comedian usually projects news from mainstream media and frames them to suit their program (Collins et al. 2020).

Propaganda Propaganda refers to news stories created by political entities to mislead people. It is a special instance of fabricated stories that aim to harm the interests of a particular party and, typically, has a political context (Zannettou et al. 2019). Propaganda was widely used during both World Wars (Collins et al. 2020) and during the Cold War (Zannettou et al. 2019). It is a consequential type of false information as it can change the course of human history (e.g., by changing the outcome of an election) (Zannettou et al. 2019). States are the main actors of propaganda. Recently, propaganda has been used by politicians and media organizations to support a certain position or view (Collins et al. 2020). Online astroturfing can be an example of the tools used for the dissemination of propaganda. It is a covert manipulation of public opinion (Peng et al. 2017) that aims to make it seem that many people share the same opinion about something. Astroturfing can affect different domains of interest, based on which online astroturfing can be mainly divided into political astroturfing, corporate astroturfing and astroturfing in e-commerce or online services (Mahbub et al. 2019). Propaganda types of fake news can be debunked with manual fact-based detection models such as the use of expert-based fact-checkers (Collins et al. 2020).

Framing Framing refers to employing some aspect of reality to make content more visible, while the truth is concealed (Collins et al. 2020) to deceive and misguide readers. People will understand certain concepts based on the way they are coined and invented. An example of framing was provided by Collins et al. (2020): "suppose a leader X says "I will neutralize my opponent" simply meaning he will beat his opponent in a given election. Such a statement will be framed such as "leader X threatens to kill Y" and this framed statement provides a total misrepresentation of the original meaning.

Conspiracy Theories Conspiracy theories refer to the belief that an event is the result of secret plots generated by powerful conspirators. Conspiracy belief refers to people's adoption and belief of conspiracy theories, and it is associated with psychological, political and social factors (Douglas et al. 2019). Conspiracy theories are widespread in contemporary democracies (Sutton and Douglas 2020), and they have major consequences. For instance, lately and during the COVID-19 pandemic, conspiracy theories have been discussed from a public health perspective (Meese et al. 2020; Allington et al. 2020; Freeman et al. 2020).

4.2.3 Comparison Between Most Popular Intent-based Types of Fake News

Following a review of the most popular intent-based types of fake news, we compare them as shown in Table 5 based on the most common criteria mentioned by researchers in their definitions as listed below.

- the intent behind the news, which refers to whether a given news type was mainly created to intentionally deceive people or not (e.g., humor, irony, entertainment, etc.);
- the way that the news propagates through OSN, which determines the nature of the propagation of each type of fake news and this can be either fast or slow propagation;
- the severity of the impact of the news on OSN users, which refers to whether the public has been highly impacted by the given type of fake news; the mentioned impact of each fake news type is mainly the proportion of the negative impact;
- and the goal behind disseminating the news, which can be to gain popularity for a particular entity (e.g., political party), for profit (e.g., lucrative business), or other reasons such as humor and irony in the case of satire, spreading panic or anger, and manipulating the public in the case of hoaxes, made-up stories about a particular person or entity in the case of rumors, and misguiding readers in the case of framing.

However, the comparison provided in Table 5 is deduced from the studied research papers; it is our point of view, which is not based on empirical data.

We suspect that the most dangerous types of fake news are the ones with high intention to deceive the public, fast

propagation through social media, high negative impact on OSN users, and complicated hidden goals and agendas. However, while the other types of fake news are less dangerous, they should not be ignored.

Moreover, it is important to highlight that the existence of the overlap in the types of fake news mentioned above has been proven, thus it is possible to observe false information that may fall within multiple categories (Zannettou et al. 2019). Here, we provide two examples by Zannettou et al. (2019) to better understand possible overlaps: (1) a rumor may also use clickbait techniques to increase the audience that will read the story; and (2) propaganda stories, as a special instance of a framing story.

5 Challenges related to fake news detection and mitigation

To alleviate fake news and its threats, it is crucial to first identify and understand the factors involved that continue to challenge researchers. Thus, the main question is to explore and investigate the factors that make it easier to fall for manipulated information. Despite the tremendous progress made in alleviating some of the challenges in fake news detection (Sharma et al. 2019; Zhou and Zafarani 2020; Zhang and Ghorbani 2020; Shu et al. 2020a), much more work needs to be accomplished to address the problem effectively.

In this section, we discuss several open issues that have been making fake news detection in social media a challenging problem. These issues can be summarized as follows: content-based issues (i.e., deceptive content that resembles the truth very closely), contextual issues (i.e., lack of user awareness, social bots spreaders of fake content, and OSN's dynamic natures that leads to the fast propagation), as well as the issue of existing datasets (i.e., there still no one size fits all benchmark dataset for fake news detection). These various aspects have proven (Shu et al. 2017) to have a great impact on the accuracy of fake news detection approaches.

5.1 Content-based issue, deceptive content

Automatic fake news detection remains a huge challenge, primarily because the content is designed in a way that it closely resembles the truth. Besides, most deceivers choose their words carefully and use their language strategically to avoid being caught. Therefore, it is often hard to determine its veracity by AI without the reliance on additional information from third parties such as fact-checkers.

Abdullah-All-Tanvir et al. (2020) reported that fake news tends to have more complicated stories and hardly ever make any references. It is more likely to contain a greater number of words that express negative emotions. This makes it so complicated that it becomes impossible for a human to manually detect the credibility of this content. Therefore, detecting fake news on social media is quite challenging. Moreover, fake news appears in multiple types and forms, which makes it hard and challenging to define a single global solution able to capture and deal with the disseminated content. Consequently, detecting false information is not a straightforward task due to its various types and forms Zannettou et al. (2019).

5.2 Contextual issues

Contextual issues are challenges that we suspect may not be related to the content of the news but rather they are inferred from the context of the online news post (i.e., humans are the weakest factor due to lack of user awareness, social bots spreaders, dynamic nature of online social platforms and fast propagation of fake news).

5.2.1 Humans are the weakest factor due to the lack of awareness

Recent statistics[31] show that the percentage of unintentional fake news spreaders (people who share fake news without the intention to mislead) over social media is five times higher than intentional spreaders. Moreover, another recent statistic[32] shows that the percentage of people who were confident about their ability to discern fact from fiction is ten times higher than those who were not confident about the truthfulness of what they are sharing. As a result, we can deduce the lack of human awareness about the ascent of fake news.

Public susceptibility and lack of user awareness (Sharma et al. 2019) have always been the most challenging problem when dealing with fake news and misinformation. This is a complex issue because many people believe almost everything on the Internet and the ones who are new to digital technology or have less expertise may be easily fooled (Edgerly et al. 2020).

Moreover, it has been widely proven (Metzger et al. 2020; Edgerly et al. 2020) that people are often motivated to support and accept information that goes with their preexisting viewpoints and beliefs, and reject information that does not fit in as well. Hence, Shu et al. (2017) illustrate an interesting correlation between fake news spread and psychological and cognitive theories. They further suggest that humans are more likely to believe information that confirms their existing views and ideological beliefs. Consequently, they deduce

[31] https://www.statista.com/statistics/657111/fake-news-sharing-online/, last access date: 26-12-2022.

[32] https://www.statista.com/statistics/657090/fake-news-recogition-confidence/, last access date: 26-12-2022.

that humans are naturally not very good at differentiating real information from fake information.

Recent research by Giachanou et al. (2020) studies the role of personality and linguistic patterns in discriminating between fake news spreaders and fact-checkers. They classify a user as a potential fact-checker or a potential fake news spreader based on features that represent users' personality traits and linguistic patterns used in their tweets. They show that leveraging personality traits and linguistic patterns can improve the performance in differentiating between checkers and spreaders.

Furthermore, several researchers studied the prevalence of fake news on social networks during (Allcott and Gentzkow 2017; Grinberg et al. 2019; Guess et al. 2019; Baptista and Gradim 2020) and after (Garrett and Bond 2021) the 2016 US presidential election and found that individuals most likely to engage with fake news sources were generally conservative-leaning, older, and highly engaged with political news.

Metzger et al. (2020) examine how individuals evaluate the credibility of biased news sources and stories. They investigate the role of both cognitive dissonance and credibility perceptions in selective exposure to attitude-consistent news information. They found that online news consumers tend to perceive attitude-consistent news stories as more accurate and more credible than attitude-inconsistent stories.

Similarly, Edgerly et al. (2020) explore the impact of news headlines on the audience's intent to verify whether given news is true or false. They concluded that participants exhibit higher intent to verify the news only when they believe the headline to be true, which is predicted by perceived congruence with preexisting ideological tendencies.

Luo et al. (2022) evaluate the effects of endorsement cues in social media on message credibility and detection accuracy. Results showed that headlines associated with a high number of likes increased credibility, thereby enhancing detection accuracy for real news but undermining accuracy for fake news. Consequently, they highlight the urgency of empowering individuals to assess both news veracity and endorsement cues appropriately on social media.

Moreover, misinformed people are a greater problem than uninformed people (Kuklinski et al. 2000), because the former hold inaccurate opinions (which may concern politics, climate change, medicine) that are harder to correct. Indeed, people find it difficult to update their misinformation-based beliefs even after they have been proved to be false (Flynn et al. 2017). Moreover, even if a person has accepted the corrected information, his/her belief may still affect their opinion (Nyhan and Reifler 2015).

Falling for disinformation may also be explained by a lack of critical thinking and of the need for evidence that supports information (Vilmer et al. 2018; Badawy et al. 2019). However, it is also possible that people choose misinformation because they engage in directionally motivated reasoning (Badawy et al. 2019; Flynn et al. 2017). Online clients are normally vulnerable and will, in general, perceive web-based networking media as reliable, as reported by Abdullah-All-Tanvir et al. (2019), who propose to mechanize fake news recognition.

It is worth noting that in addition to bots causing the outpouring of the majority of the misrepresentations, specific individuals are also contributing a large share of this issue (Abdullah-All-Tanvir et al. 2019). Furthermore, Vosoughi et al. (Vosoughi et al. 2018) found that contrary to conventional wisdom, robots have accelerated the spread of real and fake news at the same rate, implying that fake news spreads more than the truth because humans, not robots, are more likely to spread it.

In this case, verified users and those with numerous followers were not necessarily responsible for spreading misinformation of the corrupted posts (Abdullah-All-Tanvir et al. 2019).

Viral fake news can cause much havoc to our society. Therefore, to mitigate the negative impact of fake news, it is important to analyze the factors that lead people to fall for misinformation and to further understand why people spread fake news (Cheng et al. 2020). Measuring the accuracy, credibility, veracity and validity of news contents can also be a key countermeasure to consider.

5.2.2 Social bots spreaders

Several authors (Shu et al. 2018b, 2017; Shi et al. 2019; Bessi and Ferrara 2016; Shao et al. 2018a) have also shown that fake news is likely to be created and spread by non-human accounts with similar attributes and structure in the network, such as social bots (Ferrara et al. 2016). Bots (short for software robots) exist since the early days of computers. A social bot is a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior (Ferrara et al. 2016). Although they are designed to provide a useful service, they can be harmful, for example when they contribute to the spread of unverified information or rumors (Ferrara et al. 2016). However, it is important to note that bots are simply tools created and maintained by humans for some specific hidden agendas.

Social bots tend to connect with legitimate users instead of other bots. They try to act like a human with fewer words and fewer followers on social media. This contributes to the forwarding of fake news (Jiang et al. 2019). Moreover, there is a difference between bot-generated and human-written clickbait (Le et al. 2019).

Many researchers have addressed ways of identifying and analyzing possible sources of fake news spread in social media. Recent research by Shu et al. (2020a) describes

social bots use of two strategies to spread low-credibility content. First, they amplify interactions with content as soon as it is created to make it look legitimate and to facilitate its spread across social networks. Next, they try to increase public exposure to the created content and thus boost its perceived credibility by targeting influential users that are more likely to believe disinformation in the hope of getting them to "repost" the fabricated content. They further discuss the social bot detection systems taxonomy proposed by Ferrara et al. (2016) which divides bot detection methods into three classes: (1) graph-based, (2) crowdsourcing and (3) feature-based social bot detection methods.

Similarly, Shao et al. (2018a) examine social bots and how they promote the spread of misinformation through millions of Twitter posts during and following the 2016 US presidential campaign. They found that social bots played a disproportionate role in spreading articles from low-credibility sources by amplifying such content in the early spreading moments and targeting users with many followers through replies and mentions to expose them to this content and induce them to share it.

Ismailov et al. (2020) assert that the techniques used to detect bots depend on the social platform and the objective. They note that a malicious bot designed to make friends with as many accounts as possible will require a different detection approach than a bot designed to repeatedly post links to malicious websites. Therefore, they identify two models for detecting malicious accounts, each using a different set of features. Social context models achieve detection by examining features related to an account's social presence including features such as relationships to other accounts, similarities to other users' behaviors, and a variety of graph-based features. User behavior models primarily focus on features related to an individual user's behavior, such as frequency of activities (e.g., number of tweets or posts per time interval), patterns of activity and clickstream sequences.

Therefore, it is crucial to consider bot detection techniques to distinguish bots from normal users to better leverage user profile features to detect fake news.

However, there is also another "bot-like" strategy that aims to massively promote disinformation and fake content in social platforms, which is called bot farms or also troll farms. It is not social bots, but it is a group of organized individuals engaging in trolling or bot-like promotion of narratives in a coordinated fashion (Wardle 2018) hired to massively spread fake news or any other harmful content. A prominent troll farm example is the Russia-based Internet Research Agency (IRA), which disseminated inflammatory content online to influence the outcome of the 2016 U.S. presidential election.[33] As a result, Twitter suspended accounts connected to the IRA and deleted 200,000 tweets from Russian trolls (Jamieson 2020). Another example to mention in this category is review bombing (Moro and Birt 2022). Review bombing refers to coordinated groups of people massively performing the same negative actions online (e.g., dislike, negative review/comment) on an online video, game, post, product, etc., in order to reduce its aggregate review score. The review bombers can be both humans and bots coordinated in order to cause harm and mislead people by falsifying facts.

5.2.3 Dynamic nature of online social platforms and fast propagation of fake news

Sharma et al. (2019) affirm that the fast proliferation of fake news through social networks makes it hard and challenging to assess the information's credibility on social media. Similarly, Qian et al. (2018) assert that fake news and fabricated content propagate exponentially at the early stage of its creation and can cause a significant loss in a short amount of time (Friggeri et al. 2014) including manipulating the outcome of political events (Liu and Wu 2018; Bessi and Ferrara 2016).

Moreover, while analyzing the way source and promoters of fake news operate over the web through multiple online platforms, Zannettou et al. (2019) discovered that false information is more likely to spread across platforms (18% appearing on multiple platforms) compared to real information (11%).

Furthermore, recently, Shu et al. (2020c) attempted to understand the propagation of disinformation and fake news in social media and found that such content is produced and disseminated faster and easier through social media because of the low barriers that prevent doing so. Similarly, Shu et al. (2020b) studied hierarchical propagation networks for fake news detection. They performed a comparative analysis between fake and real news from structural, temporal and linguistic perspectives. They demonstrated the potential of using these features to detect fake news and they showed their effectiveness for fake news detection as well.

Lastly, Abdullah-All-Tanvir et al. (2020) note that it is almost impossible to manually detect the sources and authenticity of fake news effectively and efficiently, due to its fast circulation in such a small amount of time. Therefore, it is crucial to note that the dynamic nature of the various online social platforms, which results in the continued rapid and exponential propagation of such fake content, remains a major challenge that requires further investigation while defining innovative solutions for fake news detection.

[33] https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731, last access date: 26-12-2022.

5.3 Datasets issue

The existing approaches lack an inclusive dataset with derived multidimensional information to detect fake news characteristics to achieve higher accuracy of machine learning classification model performance (Nyow and Chua 2019). These datasets are primarily dedicated to validating the machine learning model and are the ultimate frame of reference to train the model and analyze its performance. Therefore, if a researcher evaluates their model based on an unrepresentative dataset, the validity and the efficiency of the model become questionable when it comes to applying the fake news detection approach in a real-world scenario.

Moreover, several researchers (Shu et al. 2020d; Wang et al. 2020; Pathak and Srihari 2019; Przybyla 2020) believe that fake news is diverse and dynamic in terms of content, topics, publishing methods and media platforms, and sophisticated linguistic styles geared to emulate true news. Consequently, training machine learning models on such sophisticated content requires large-scale annotated fake news data that are difficult to obtain (Shu et al. 2020d).

Therefore, datasets are also a great topic to work on to enhance data quality and have better results while defining our solutions. Adversarial learning techniques (e.g., GAN, SeqGAN) can be used to provide machine-generated data that can be used to train deeper models and build robust systems to detect fake examples from the real ones. This approach can be used to counter the lack of datasets and the scarcity of data available to train models.

6 Fake news detection literature review

Fake news detection in social networks is still in the early stage of development and there are still challenging issues that need further investigation. This has become an emerging research area that is attracting huge attention.

There are various research studies on fake news detection in online social networks. Few of them have focused on the automatic detection of fake news using artificial intelligence techniques. In this section, we review the existing approaches used in automatic fake news detection, as well as the techniques that have been adopted. Then, a critical discussion built on a primary classification scheme based on a specific set of criteria is also emphasized.

6.1 Categories of fake news detection

In this section, we give an overview of most of the existing automatic fake news detection solutions adopted in the literature. A recent classification by Sharma et al. (2019) uses three categories of fake news identification methods. Each category is further divided based on the type of existing methods (i.e., content-based, feedback-based and intervention-based methods). However, a review of the literature for fake news detection in online social networks shows that the existing studies can be classified into broader categories based on two major aspects that most authors inspect and make use of to define an adequate solution. These aspects can be considered as major sources of extracted information used for fake news detection and can be summarized as follows: the content-based (i.e., related to the content of the news post) and the contextual aspect (i.e., related to the context of the news post).

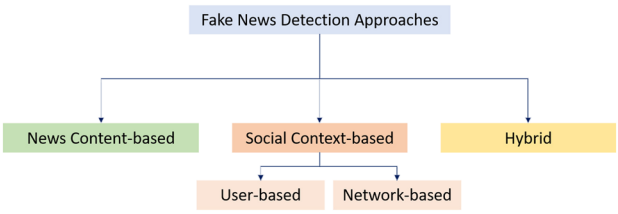

Fig. 5 Classification of fake news detection approaches

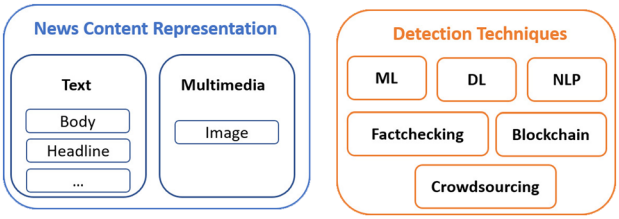

Fig. 6 News content-based category: news content representation and detection techniques

Consequently, the studies we reviewed can be classified into three different categories based on the two aspects mentioned above (the third category is hybrid). As depicted in Fig. 5, fake news detection solutions can be categorized as news content-based approaches, the social context-based approaches that can be divided into network and user-based approaches, and hybrid approaches. The latter combines both content-based and contextual approaches to define the solution.

6.1.1 News Content-based Category

News content-based approaches are fake news detection approaches that use content information (i.e., information extracted from the content of the news post) and that focus on studying and exploiting the news content in their proposed solutions. Content refers to the body of the news, including source, headline, text and image-video, which can reflect subtle differences.

Researchers of this category rely on content-based detection cues (i.e., text and multimedia-based cues), which are

features extracted from the content of the news post. Text-based cues are features extracted from the text of the news, whereas multimedia-based cues are features extracted from the images and videos attached to the news. Figure 6 summarizes the most widely used news content representation (i.e., text and multimedia/images) and detection techniques (i.e., machine learning (ML), deep Learning (DL), natural language processing (NLP), fact-checking, crowdsourcing (CDS) and blockchain (BKC)) in news content-based category of fake news detection approaches. Most of the reviewed research works based on news content for fake news detection rely on the text-based cues (Kapusta et al. 2019; Kaur et al. 2020; Vereshchaka et al. 2020; Ozbay and Alatas 2020; Wang 2017; Nyow and Chua 2019; Hosseinimotlagh and Papalexakis 2018; Abdullah-All-Tanvir et al. 2019, 2020; Mahabub 2020; Bahad et al. 2019; Hiriyannaiah et al. 2020) extracted from the text of the news content including the body of the news and its headline. However, a few researchers such as Vishwakarma et al. (2019) and Amri et al. (2022) try to recognize text from the associated image.

Most researchers of this category rely on artificial intelligence (AI) techniques (such as ML, DL and NLP models) to improve performance in terms of prediction accuracy. Others use different techniques such as fact-checking, crowdsourcing and blockchain. Specifically, the AI- and ML-based approaches in this category are trying to extract features from the news content, which they use later for content analysis and training tasks. In this particular case, the extracted features are the different types of information considered to be relevant for the analysis. Feature extraction is considered as one of the best techniques to reduce data size in automatic fake news detection. This technique aims to choose a subset of features from the original set to improve classification performance (Yazdi et al. 2020).

Table 6 lists the distinct features and metadata, as well as the used datasets in the news content-based category of fake news detection approaches.

6.1.2 Social Context-based Category

Unlike news content-based solutions, the social context-based approaches capture the skeptical social context of the online news (Zhang and Ghorbani 2020) rather than focusing on the news content. The social context-based category contains fake news detection approaches that use the contextual aspects (i.e., information related to the context of the news post). These aspects are based on social context and they offer additional information to help detect fake news. They are the surrounding data outside of the fake news article itself, where they can be an essential part of automatic fake news detection. Some useful examples of contextual information may include checking if the news itself and the source that published it are credible, checking the date of the news or the supporting resources, and checking if any other online news platforms are reporting the same or similar stories (Zhang and Ghorbani 2020).

Social context-based aspects can be classified into two subcategories, user-based and network-based, and they can be used for context analysis and training tasks in the case of AI- and ML-based approaches. User-based aspects refer to information captured from OSN users such as user profile information (Shu et al. 2019b; Wang et al. 2019c; Hamdi et al. 2020; Nyow and Chua 2019; Jiang et al. 2019) and user behavior (Cardaioli et al. 2020) such as user engagement (Uppada et al. 2022; Jiang et al. 2019; Shu et al. 2018b; Nyow and Chua 2019) and response (Zhang et al. 2019a; Qian et al. 2018). Meanwhile, network-based aspects refer to information captured from the properties of the social network where the fake content is shared and disseminated such as news propagation path (Liu and Wu 2018; Wu and Liu 2018) (e.g., propagation times and temporal characteristics of propagation), diffusion patterns (Shu et al. 2019a) (e.g., number of retweets, shares), as well as user relationships (Mishra 2020; Hamdi et al. 2020; Jiang et al. 2019) (e.g., friendship status among users).

Figure 7 summarizes some of the most widely adopted social context representations, as well as the most used detection techniques (i.e., AI, ML, DL, fact-checking and blockchain), in the social context-based category of approaches.

Table 7 lists the distinct features and metadata, the adopted detection cues, as well as the used datasets, in the context-based category of fake news detection approaches.

6.1.3 Hybrid approaches

Most researchers are focusing on employing a specific method rather than a combination of both content- and context-based methods. This is because some of them (Wu and Rao 2020) believe that there still some challenging limitations in the traditional fusion strategies due to existing feature correlations and semantic conflicts. For this reason, some researchers focus on extracting content-based information, while others are capturing some social context-based information for their proposed approaches.

However, it has proven challenging to successfully automate fake news detection based on just a single type of feature (Ruchansky et al. 2017). Therefore, recent directions tend to do a mixture by using both news content-based and social context-based approaches for fake news detection.

Table 8 lists the distinct features and metadata, as well as the used datasets, in the hybrid category of fake news detection approaches.

Table 6 The features and datasets used in the news content-based approaches

Feature and metadata	Datasets	Reference
The average number of words in sentences, number of stop words, the sentiment rate of the news measured through the difference between the number of positive and negative words in the article	Getting real about fake news[a], Gathering media-biasfactcheck[b], KaiDMML FakeNewsNet[c], Real news for Oct-Dec 2016[d]	Kapusta et al. (2019)
The length distribution of the title, body and label of the article	News trends, Kaggle, Reuters	Kaur et al. (2020)
Sociolinguistic, historical, cultural, ideological and syntactical features attached to particular words, phrases and syntactical constructions	FakeNewsNet	Vereshchaka et al. (2020)
Term frequency	BuzzFeed political news, Random political news, ISOT fake news	Ozbay and Alatas (2020)
The statement, speaker, context, label, justification	POLITIFACT, LIAR[e]	Wang (2017)
Spatial vicinity of each word, spatial/contextual relations between terms, and latent relations between terms and articles	Kaggle fake news dataset[f]	Hosseinimotlagh and Papalexakis (2018)
Word length, the count of words in a tweeted statement	Twitter dataset, Chile earthquake 2010 datasets	Abdullah-All-Tanvir et al. (2019)
The number of words that express negative emotions	Twitter dataset	Abdullah-All-Tanvir et al. (2020)
Labeled data	BuzzFeed[g], PolitiFact[h]	Mahabub (2020)
The relationship between the news article headline and article body. The biases of a written news article	Kaggle: real_or_fake[i], Fake news detection[j]	Bahad et al. (2019)
Historical data. The topic and sentiment associated with content textual. The subject and context of the text, semantic knowledge of the content	Facebook dataset	Del Vicario et al. (2019)
The veracity of image text. The credibility of the top 15 Google search results related to the image text	Google images, the Onion, Kaggle	Vishwakarma et al. (2019)
Topic modeling of text and the associated image of the online news	Twitter dataset[k], Weibo[l]	Amri et al. (2022)

[a] https://www.kaggle.com/anthonyc1/gathering-real-news-for-oct-dec-2016, last access date: 26-12-2022

[b] https://mediabiasfactcheck.com/, last access date: 26-12-2022

[c] https://github.com/KaiDMML/FakeNewsNet, last access date: 26-12-2022

[d] https://www.kaggle.com/anthonyc1/gathering-real-news-for-oct-dec-2016, last access date: 26-12-2022

[e] https://www.cs.ucsb.edu/~william/data/liar_dataset.zip, last access date: 26-12-2022

[f] https://www.kaggle.com/mrisdal/fake-news, last access date: 26-12-2022

[g] https://github.com/BuzzFeedNews/2016-10-facebook-fact-check, last access date: 26-12-2022

[h] https://www.politifact.com/subjects/fake-news/, last access date: 26-12-2022

[i] https://www.kaggle.com/rchitic17/real-or-fake, last access date: 26-12-2022

[j] https://www.kaggle.com/jruvika/fake-news-detection, last access date: 26-12-2022

[k] https://github.com/MKLab-ITI/image-verification-corpus, last access date: 26-12-2022

[l] https://drive.google.com/file/d/14VQ7EWPiFeGzxp3XC2DeEHi-BEisDINn/view, last access date: 26-12-2022

6.2 Fake news detection techniques

Another vision for classifying automatic fake news detection is to look at techniques used in the literature. Hence, we classify the detection methods based on the techniques into three groups:

- Human-based techniques: This category mainly includes the use of crowdsourcing and fact-checking techniques, which rely on human knowledge to check and validate the veracity of news content.
- Artificial Intelligence-based techniques: This category includes the most used AI approaches for fake news

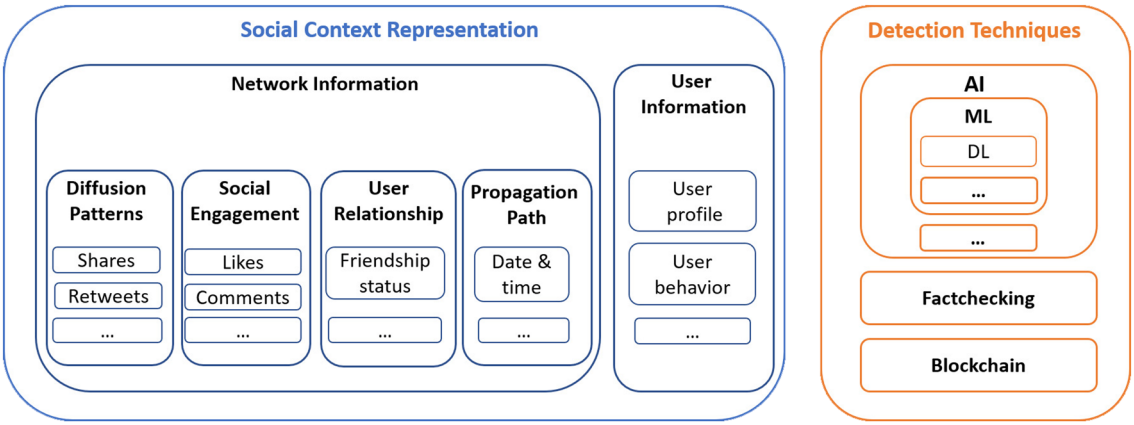

Fig. 7 Social context-based category: social context representation and detection techniques

detection in the literature. Specifically, these are the approaches in which researchers use classical ML, deep learning techniques such as convolutional neural network (CNN), recurrent neural network (RNN), as well as natural language processing (NLP).

– Blockchain-based techniques: This category includes solutions using blockchain technology to detect and mitigate fake news in social media by checking source reliability and establishing the traceability of the news content.

6.2.1 Human-based Techniques

One specific research direction for fake news detection consists of using human-based techniques such as crowdsourcing (Pennycook and Rand 2019; Micallef et al. 2020) and fact-checking (Vlachos and Riedel 2014; Chung and Kim 2021; Nyhan et al. 2020) techniques.

These approaches can be considered as low computational requirement techniques since both rely on human knowledge and expertise for fake news detection. However, fake news identification cannot be addressed solely through human force since it demands a lot of effort in terms of time and cost, and it is ineffective in terms of preventing the fast spread of fake content.

Crowdsourcing. Crowdsourcing approaches (Kim et al. 2018) are based on the "wisdom of the crowds" (Collins et al. 2020) for fake content detection. These approaches rely on the collective contributions and crowd signals (Tschiatschek et al. 2018) of a group of people for the aggregation of crowd intelligence to detect fake news (Tchakounté et al. 2020) and to reduce the spread of misinformation on social media (Pennycook and Rand 2019; Micallef et al. 2020).

Micallef et al. (2020) highlight the role of the crowd in countering misinformation. They suspect that concerned citizens (i.e., the crowd), who use platforms where disinformation appears, can play a crucial role in spreading fact-checking information and in combating the spread of misinformation.

Recently Tchakounté et al. (2020) proposed a voting system as a new method of binary aggregation of opinions of the crowd and the knowledge of a third-party expert. The aggregator is based on majority voting on the crowd side and weighted averaging on the third-party site.

Similarly, Huffaker et al. (2020) propose a crowdsourced detection of emotionally manipulative language. They introduce an approach that transforms classification problems into a comparison task to mitigate conflation content by allowing the crowd to detect text that uses manipulative emotional language to sway users toward positions or actions. The proposed system leverages anchor comparison to distinguish between intrinsically emotional content and emotionally manipulative language.

La Barbera et al. (2020) try to understand how people perceive the truthfulness of information presented to them. They collect data from US-based crowd workers, build a dataset of crowdsourced truthfulness judgments for political statements, and compare it with expert annotation data generated by fact-checkers such as PolitiFact.

Coscia and Rossi (2020) introduce a crowdsourced flagging system that consists of online news flagging. The bipolar model of news flagging attempts to capture the main ingredients that they observe in empirical research on fake news and disinformation.

Unlike the previously mentioned researchers who focus on news content in their approaches, Pennycook and Rand (2019) focus on using crowdsourced judgments of the quality of news sources to combat social media disinformation.

Fact-Checking. The fact-checking task is commonly manually performed by journalists to verify the truthfulness of a given claim. Indeed, fact-checking features are being adopted by multiple online social network platforms. For

Table 7 The features, detection cues and datasets used int the social context-based approaches

Feature and metadata	Detection cues	Datasets	Reference
Users' sharing behaviors, explicit and implicit profile features	User-based: user profile information	FakeNewsNet	Shu et al. (2019b)
Users' trust level, explicit and implicit profile features of "experienced" users who can recognize fake news items as false and "naive" users who are more likely to believe fake news	User-based: user engagement	FakeNewsNet, BuzzFeed, PolitiFact	Shu et al. (2018b)
Users' replies on fake content, the reply stances	User-based: user response	RumourEval, PHEME	Zhang et al. (2019a)
Historical user responses to previous articles	User-based: user response	Weibo, Twitter dataset	Qian et al. (2018)
Speaker name, job title, political party affiliation, etc.	User-based: user profile information	LIAR	Wang et al. (2019b)
Latent relationships among users, the influence of the users with high prestige on the other users	Networks-based: user relationships	Twitter15 and Twitter16[a]	Mishra (2020)
The inherent tri-relationships among publishers, news items and users (i.e., publisher-news relations and user-news interactions)	Networks-based: diffusion patterns	FakeNewsNet	Shu et al. (2019b)
Propagation paths of news stories constructed from the retweets of source tweets	Networks-based: news propagation path	Weibo, Twitter15, Twitter16	Liu and Wu (2018)
The propagation of messages in a social network	Networks-based: news propagation path	Twitter dataset	Wu and Liu (2018)
Spatiotemporal information (i.e., location, timestamps of user engagements), user's Twitter profile, the user engagement to both fake and real news	User-based: user engagement	FakeNewsNet, PolitiFact, GossipCop, Twitter	Nyow and Chua (2019)
The credibility of information sources, characteristics of the user, and their social graph	User and network-based: user profile information and user relationships	Ego-Twitter[b]	Hamdi et al. (2020)
Number of follows and followers on social media (user followee/follower, The friendship network), users' similarities	User and network-based: user profile information, user engagement and user relationships	FakeNewsNet	Jiang et al. (2019)

[a] https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip, last access date: 26-12-2022[b] https://snap.stanford.edu/data/ego-Twitter.html, last access date: 26-12-2022

Table 8 The features and datasets used in the hybrid approaches

Feature and metadata	Datasets	Reference
Features and textual metadata of the news content: title, content, date, source, location	SOT fake news dataset, LIAR dataset and FA-KES dataset	Elhadad et al. (2019)
Spatiotemporal information (i.e., location, timestamps of user engagements), user's Twitter profile, the user engagement to both fake and real news	FakeNewsNet, PolitiFact, GossipCop, Twitter	Nyow and Chua (2019)
The domains and reputations of the news publishers. The important terms of each news and their word embeddings and topics. Shares, reactions and comments	BuzzFeed	Xu et al. (2019)
Shares and propagation path of the tweeted content. A set of metrics comprising of created discussions such as the increase in authors, attention level, burstiness level, contribution sparseness, author interaction, author count and the average length of discussions	Twitter dataset	Aswani et al. (2017)
Features extracted from the evolution of news and features from the users involved in the news spreading: The news veracity, the credibility of news spreaders, and the frequency of exposure to the same piece of news	Twitter dataset	Previti et al. (2020)
Similar semantics and conflicting semantics between posts and comments	RumourEval, PHEME	Wu and Rao (2020)
Information from the publisher, including semantic and emotional information in news content. Semantic and emotional information from users. The resultant latent representations from news content and user comments	Weibo	Guo et al. (2019)
Relationships between news articles, creators and subjects	PolitiFact	Zhang et al. (2020)
Source domains of the news article, author names	George McIntire fake news dataset	Deepak and Chitturi (2020)
The news content, social context and spatiotemporal information. Synthetic user engagements generated from historical temporal user engagement patterns	FakeNewsNet	Shu et al. (2018a)
The news content, social reactions, statements, the content and language of posts, the sharing and dissemination among users, content similarity, stance, sentiment score, headline, named entity, news sharing, credibility history, tweet comments	SHPT, PolitiFact	Wang et al. (2019a)
The source of the news, its headline, its author, its publication time, the adherence of a news source to a particular party, likes, shares, replies, followers-followees and their activities	NELA-GT-2019, Fakeddit	Raza and Ding (2022)

instance, Facebook[34] started addressing false information through independent fact-checkers in 2017, followed by Google[35] the same year. Two years later, Instagram[36] followed suit. However, the usefulness of fact-checking initiatives is questioned by journalists[37], as well as by researchers such as Andersen and Søe (2020). On the other hand, work is being conducted to boost the effectiveness of these initiatives to reduce misinformation (Chung and Kim 2021; Clayton et al. 2020; Nyhan et al. 2020).

Most researchers use fact-checking websites (e.g., politifact.com,[38] snopes.com,[39] Reuters,[40], etc.) as data sources to build their datasets and train their models. Therefore, in the following, we specifically review examples of solutions that use fact-checking (Vlachos and Riedel 2014) to help build datasets that can be further used in the automatic detection of fake content.

Yang et al. (2019a) use PolitiFact fact-checking website as a data source to train, tune, and evaluate their model named XFake, on political data. The XFake system is an explainable fake news detector that assists end users to identify news credibility. The fakeness of news items is detected

[34] https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news, last access date: 26-12-2022.

[35] https://www.theguardian.com/technology/2017/apr/07/google-to-display-fact-checking-labels-to-show-if-news-is-true-or-false, last access date: 26-12-2022.

[36] https://about.instagram.com/blog/announcements/combatting-misinformation-on-instagram, last access date: 26-12-2022.

[37] https://www.wired.com/story/instagram-fact-checks-who-will-do-checking/, last access date: 26-12-2022.

[38] https://www.politifact.com/, last access date: 26-12-2022.

[39] https://www.snopes.com/, last access date: 26-12-2022.

[40] https://www.reutersagency.com/en/, last access date: 26-12-2022.

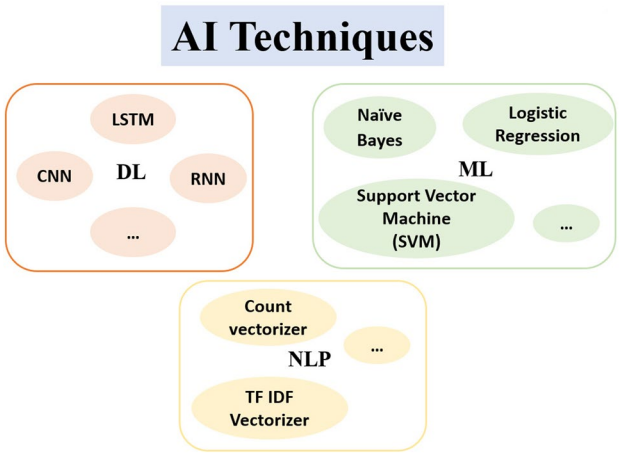

Fig. 8 Examples of the most widely used AI techniques for fake news detection

and interpreted considering both content and contextual (e.g., statements) information (e.g., speaker).

Based on the idea that fact-checkers cannot clean all data, and it must be a selection of what "matters the most" to clean while checking a claim, Sintos et al. (2019) propose a solution to help fact-checkers combat problems related to data quality (where inaccurate data lead to incorrect conclusions) and data phishing. The proposed solution is a combination of data cleaning and perturbation analysis to avoid uncertainties and errors in data and the possibility that data can be phished.

Tchechmedjiev et al. (2019) propose a system named "ClaimsKG" as a knowledge graph of fact-checked claims aiming to facilitate structured queries about their truth values, authors, dates, journalistic reviews and other kinds of metadata. "ClaimsKG" designs the relationship between vocabularies. To gather vocabularies, a semi-automated pipeline periodically gathers data from popular fact-checking websites regularly.

6.2.2 AI-based Techniques

Previous work by Yaqub et al. (2020) has shown that people lack trust in automated solutions for fake news detection However, work is already being undertaken to increase this trust, for instance by von der Weth et al. (2020).

Most researchers consider fake news detection as a classification problem and use artificial intelligence techniques, as shown in Fig. 8. The adopted AI techniques may include machine learning ML (e.g., Naïve Bayes, logistic regression, support vector machine SVM), deep learning DL (e.g., convolutional neural networks CNN, recurrent neural networks RNN, long short-term memory LSTM) and natural language processing NLP (e.g., Count vectorizer, TF-IDF Vectorizer). Most of them combine many AI techniques in their solutions rather than relying on one specific approach.

Many researchers are developing machine learning models in their solutions for fake news detection. Recently, deep neural network techniques are also being employed as they are generating promising results (Islam et al. 2020). A neural network is a massively parallel distributed processor with simple units that can store important information and make it available for use (Hiriyannaiah et al. 2020). Moreover, it has been proven (Cardoso Durier da Silva et al. 2019) that the most widely used method for automatic detection of fake news is not simply a classical machine learning technique, but rather a fusion of classical techniques coordinated by a neural network.

Some researchers define purely machine learning models (Del Vicario et al. 2019; Elhadad et al. 2019; Aswani et al. 2017; Hakak et al. 2021; Singh et al. 2021) in their fake news detection approaches. The more commonly used machine learning algorithms (Abdullah-All-Tanvir et al. 2019) for classification problems are Naïve Bayes, logistic regression and SVM.

Other researchers (Wang et al. 2019c; Wang 2017; Liu and Wu 2018; Mishra 2020; Qian et al. 2018; Zhang et al. 2020; Goldani et al. 2021) prefer to do a mixture of different deep learning models, without combining them with classical machine learning techniques. Some even prove that deep learning techniques outperform traditional machine learning techniques (Mishra et al. 2022). Deep learning is one of the most widely popular research topics in machine learning. Unlike traditional machine learning approaches, which are based on manually crafted features, deep learning approaches can learn hidden representations from simpler inputs both in context and content variations (Bondielli and Marcelloni 2019). Moreover, traditional machine learning algorithms almost always require structured data and are designed to "learn" to act by understanding labeled data and then use it to produce new results with more datasets, which requires human intervention to "teach them" when the result is incorrect (Parrish 2018), while deep learning networks rely on layers of artificial neural networks (ANN) and do not require human intervention, as multilevel layers in neural networks place data in a hierarchy of different concepts, which ultimately learn from their own mistakes (Parrish

Table 9 A classification of popular blockchain-based approaches for fake news detection in social media

Reference	Fake News Type		Techniques	Feature
	Multimedia	Text		
Shae and Tsai (2019)	✓	✓	AI	Reliability
Ochoa et al. (2019)	–	✓	Data Mining, Truth-Discovery	Reliability
Huckle and White (2017)	✓	–	Preservation Metadata	Reliability
Song et al. (2019)	–	–	–	Traceability
Shang et al. (2018)	–	–	–	Traceability
Qayyum et al. (2019)	–	–	Semantic Similarity	Reliability
Jing and Murugesan (2018)	–	–	AI	Reliability
Buccafurri et al. (2017)	–	–	Crowd-Sourcing	Reliability
Chen et al. (2018)	–	–	SIR Model	Reliability
Hasan and Salah (2019)	✓	–	–	Authenticity
Tchechmedjiev et al. (2019)	–	–	Graph theory	Reliability

2018). The two most widely implemented paradigms in deep neural networks are recurrent neural networks (RNN) and convolutional neural networks (CNN).

Still other researchers (Abdullah-All-Tanvir et al. 2019; Kaliyar et al. 2020; Zhang et al. 2019a; Deepak and Chitturi 2020; Shu et al. 2018a; Wang et al. 2019c) prefer to combine traditional machine learning and deep learning classification, models. Others combine machine learning and natural language processing techniques. A few combine deep learning models with natural language processing (Vereshchaka et al. 2020). Some other researchers (Kapusta et al. 2019; Ozbay and Alatas 2020; Ahmed et al. 2020) combine natural language processing with machine learning models. Furthermore, others (Abdullah-All-Tanvir et al. 2019; Kaur et al. 2020; Kaliyar 2018; Abdullah-All-Tanvir et al. 2020; Bahad et al. 2019) prefer to combine all the previously mentioned techniques (i.e., ML, DL and NLP) in their approaches.

Table 11, which is relegated to the Appendix (after the bibliography) because of its size, shows a comparison of the fake news detection solutions that we have reviewed based on their main approaches, the methodology that was used and the models.

6.2.3 Blockchain-based Techniques for Source Reliability and Traceability

Another research direction for detecting and mitigating fake news in social media focuses on using blockchain solutions. Blockchain technology is recently attracting researchers' attention due to the interesting features it offers. Immutability, decentralization, tamperproof, consensus, record keeping and non-repudiation of transactions are some of the key features that make blockchain technology exploitable, not just for cryptocurrencies, but also to prove the authenticity and integrity of digital assets.

However, the proposed blockchain approaches are few in number and they are fundamental and theoretical approaches. Specifically, the solutions that are currently available are still in research, prototype, and beta testing stages (DiCicco and Agarwal 2020; Tchechmedjiev et al. 2019). Furthermore, most researchers (Ochoa et al. 2019; Song et al. 2019; Shang et al. 2018; Qayyum et al. 2019; Jing and Murugesan 2018; Buccafurri et al. 2017; Chen et al. 2018) do not specify which fake news type they are mitigating in their studies. They mention news content in general, which is not adequate for innovative solutions. For that, serious implementations should be provided to prove the usefulness and feasibility of this newly developing research vision.

Table 9 shows a classification of the reviewed blockchain-based approaches. In the classification, we listed the following:

- The type of fake news that authors are trying to mitigate, which can be multimedia-based or text-based fake news.
- The techniques used for fake news mitigation, which can be either blockchain only, or blockchain combined with other techniques such as AI, Data mining, Truth-discovery, Preservation metadata, Semantic similarity,

Table 10 Fake news detection approaches classification

	Artificial Intelligence			Crowdsourcing (CDS)	Blockchain (BKC)	Fact-checking	Hybrid
	ML	DL	NLP				
Content	Del Vicario et al. (2019), Hosseinimotlagh and Papalexakis (2018), Hakak et al. (2021), Singh et al. (2021), Amri et al. (2022)	Wang (2017), Hiriyannaiah et al. (2020)	Zellers et al. (2019)	Kim et al. (2018), Tschiatschek et al. (2018), Tchakounté et al. (2020), Huffaker et al. (2020), La Barbera et al. (2020), Coscia and Rossi (2020), Micallef et al. (2020)	Song et al. (2019)	Sintos et al. (2019)	ML, DL, NLP: Abdullah-All-Tanvir et al. (2020), Kaur et al. (2020), Mahabub (2020), Bahad et al. (2019) Kaliyar (2018) ML, DL: Abdullah-All-Tanvir et al. (2019), Kaliyar et al. (2020), Deepak and Chitturi (2020) DL, NLP: Vereshchaka et al. (2020) ML, NLP: Kapusta et al. (2019), Ozbay and Alatas Ozbay and Alatas (2020), Ahmed et al. (2020) BKC, CDS: Buccafurri et al. (2017)
Context	–	Qian et al. (2018), Liu and Wu (2018), Hamdi et al. (2020), Wang et al. (2019b), Mishra (2020)	–	Pennycook and Rand (2019)	Huckle and White (2017), Shang et al. (2018)	Tchechmedjiev et al. (2019)	ML, DL: Zhang et al. (2019a), Shu et al. (2019b), Shu et al. (2018b), Wu and Liu (2018) BKC, AI: Ochoa et al. (2019) BKC, SIR: Chen et al. (2018)
Hybrid	Aswani et al. (2017), Previti et al. (2020), Elhadad et al. (2019), Nyow and Chua (2019)	Ruchansky et al. (2017), Wu and Rao (2020), Guo et al. (2019), Zhang et al. (2020)	Xu et al. (2019)	–	Qayyum et al. (2019), Hasan and Salah (2019), Tchechmedjiev et al. (2019)	Yang et al. (2019a)	ML, DL: Shu et al. (2018a), Wang et al. (2019b) BKC, AI: Shae and Tsai (2019), Jing and Murugesan (2018)

Crowdsourcing, Graph theory and SIR model (Susceptible, Infected, Recovered).
– The feature that is offered as an advantage of the given solution (e.g., Reliability, Authenticity and Traceability). Reliability is the credibility and truthfulness of the news content, which consists of proving the trustworthiness of the content. Traceability aims to trace and archive the contents. Authenticity consists of checking whether the content is real and authentic.

A checkmark (✓) in Table 9 denotes that the mentioned criterion is explicitly mentioned in the proposed solution, while the empty dash (–) cell for fake news type denotes that it depends on the case: The criterion was either not explicitly mentioned (e.g., fake news type) in the work or the classification does not apply (e.g., techniques/other).

7 Discussion

After reviewing the most relevant state of the art for automatic fake news detection, we classify them as shown in Table 10 based on the detection aspects (i.e., content-based, contextual, or hybrid aspects) and the techniques used (i.e., AI, crowdsourcing, fact-checking, blockchain or hybrid techniques). Hybrid techniques refer to solutions that simultaneously combine different techniques from previously mentioned categories (i.e., inter-hybrid methods), as well as techniques within the same class of methods (i.e., intra-hybrid methods), in order to define innovative solutions for fake news detection. A hybrid method should bring the best of both worlds. Then, we provide a discussion based on different axes.

7.1 News content-based methods

Most of the news content-based approaches consider fake news detection as a classification problem and they use AI techniques such as classical machine learning (e.g., regression, Bayesian) as well as deep learning (i.e., neural methods such as CNN and RNN). More specifically, classification of social media content is a fundamental task for social media mining, so that most existing methods regard it as a text categorization problem and mainly focus on using content features, such as words and hashtags (Wu and Liu 2018). The main challenge facing these approaches is how to extract features in a way to reduce the data used to train their models and what features are the most suitable for accurate results.

Researchers using such approaches are motivated by the fact that the news content is the main entity in the deception process, and it is a straightforward factor to analyze and use while looking for predictive clues of deception. However, detecting fake news only from the content of the news is not enough because the news is created in a strategic intentional way to mimic the truth (i.e., the content can be intentionally manipulated by the spreader to make it look like real news). Therefore, it is considered to be challenging, if not impossible, to identify useful features (Wu and Liu 2018) and consequently tell the nature of such news solely from the content.

Moreover, works that utilize only the news content for fake news detection ignore the rich information and latent user intelligence (Qian et al. 2018) stored in user responses toward previously disseminated articles. Therefore, the auxiliary information is deemed crucial for an effective fake news detection approach.

7.2 Social context-based methods

The context-based approaches explore the surrounding data outside of the news content, which can be an effective direction and has some advantages in areas where the content approaches based on text classification can run into issues. However, most existing studies implementing contextual methods mainly focus on additional information coming from users and network diffusion patterns. Moreover, from a technical perspective, they are limited to the use of sophisticated machine learning techniques for feature extraction, and they ignore the usefulness of results coming from techniques such as web search and crowdsourcing which may save much time and help in the early detection and identification of fake content.

7.3 Hybrid approaches

Hybrid approaches can simultaneously model different aspects of fake news such as the content-based aspects, as well as the contextual aspect based on both the OSN user and the OSN network patterns. However, these approaches are deemed more complex in terms of models (Bondielli and Marcelloni 2019), data availability, and the number of features. Furthermore, it remains difficult to decide which information among each category (i.e., content-based and context-based information) is most suitable and appropriate to be used to achieve accurate and precise results. Therefore, there are still very few studies belonging to this category of hybrid approaches.

7.4 Early detection

As fake news usually evolves and spreads very fast on social media, it is critical and urgent to consider early detection directions. Yet, this is a challenging task to do especially in highly dynamic platforms such as social networks. Both news content- and social context-based approaches suffer from this challenging early detection of fake news.

Although approaches that detect fake news based on content analysis face this issue less, they are still limited by the lack of information required for verification when the news is in its early stage of spread. However, approaches that detect fake news based on contextual analysis are most likely to suffer from the lack of early detection since most of them rely on information that is mostly available after the spread of fake content such as social engagement, user response, and propagation patterns. Therefore, it is crucial to consider both trusted human verification and historical data as an attempt to detect fake content during its early stage of propagation.

8 Conclusion and future directions

In this paper, we introduced the general context of the fake news problem as one of the major issues of the online deception problem in online social networks. Based on reviewing the most relevant state of the art, we summarized and classified existing definitions of fake news, as well as its related terms. We also listed various typologies and existing categorizations of fake news such as intent-based fake news including clickbait, hoax, rumor, satire, propaganda, conspiracy theories, framing as well as content-based fake news including text and multimedia-based fake news, and in the latter, we can tackle deepfake videos and GAN-generated fake images. We discussed the major challenges related to fake news detection and mitigation in social media including the deceptiveness nature of the fabricated content, the lack of human awareness in the field of fake news, the non-human spreaders issue (e.g., social bots), the dynamicity of such online platforms, which results in a fast propagation of fake content and the quality of existing datasets, which still limits the efficiency of the proposed solutions. We reviewed existing researchers' visions regarding the automatic detection of fake news based on the adopted approaches (i.e., news content-based approaches, social context-based approaches, or hybrid approaches) and the techniques that are used (i.e., artificial intelligence-based methods; crowdsourcing, fact-checking, and blockchain-based methods; and hybrid methods), then we showed a comparative study between the reviewed works. We also provided a critical discussion of the reviewed approaches based on different axes such as the adopted aspect for fake news detection (i.e., content-based, contextual, and hybrid aspects) and the early detection perspective.

To conclude, we present the main issues for combating the fake news problem that needs to be further investigated while proposing new detection approaches. We believe that to define an efficient fake news detection approach, we need to consider the following:

- Our choice of sources of information and search criteria may have introduced biases in our research. If so, it would be desirable to identify those biases and mitigate them.
- News content is the fundamental source to find clues to distinguish fake from real content. However, contextual information derived from social media users and from the network can provide useful auxiliary information to increase detection accuracy. Specifically, capturing users' characteristics and users' behavior toward shared content can be a key task for fake news detection.
- Moreover, capturing users' historical behavior, including their emotions and/or opinions toward news content, can help in the early detection and mitigation of fake news.
- Furthermore, adversarial learning techniques (e.g., GAN, SeqGAN) can be considered as a promising direction for mitigating the lack and scarcity of available datasets by providing machine-generated data that can be used to train and build robust systems to detect the fake examples from the real ones.
- Lastly, analyzing how sources and promoters of fake news operate over the web through multiple online platforms is crucial; Zannettou et al. (2019) discovered that false information is more likely to spread across platforms (18% appearing on multiple platforms) compared to valid information (11%).

Appendix: A Comparison of AI-based fake news detection techniques

This Appendix consists only in the rather long Table 11. It shows a comparison of the fake news detection solutions based on artificial intelligence that we have reviewed according to their main approaches, the methodology that was used, and the models, as explained in Sect. 6.2.2.

Table 11 Comparison of AI-based fake news detection techniques

Reference	Approach	Method	Model
Del Vicario et al. (2019)	An approach to analyze the sentiment associated with data textual content and add semantic knowledge to it	ML	Linear Regression (LIN), Logistic Regression (LOG), Support Vector Machine (SVM) with Linear Kernel, K-Nearest Neighbors (KNN), Neural Network Models (NN), Decision Trees (DT)
Elhadad et al. (2019)	An approach to select hybrid features from the textual content of the news, which they consider as blocks, without segmenting text into parts (title, content, date, source, etc.)	ML	Decision Tree, KNN, Logistic Regression, SVM, Naïve Bayes with n-gram, LSVM, Perceptron
Aswani et al. (2017)	A hybrid artificial bee colony approach to identify and segregate buzz in Twitter and analyze user-generated content (UGC) to mine useful information (content buzz/popularity)	ML	KNN with artificial bee colony optimization
Hakak et al. (2021)	An ensemble of machine learning approaches for effective feature extraction to classify fake news	ML	Decision Tree, Random Forest and Extra Tree Classifier
Singh et al. (2021)	A multimodal approach, combining text and visual analysis of online news stories to automatically detect fake news through predictive analysis to detect features most strongly associated with fake news	ML	Logistic Regression, Linear Discrimination Analysis, Quadratic Discriminant Analysis, K-Nearest Neighbors, Naïve Bayes, Support Vector Machine, Classification and Regression Tree, and Random Forest Analysis
Amri et al. (2022)	An explainable multimodal content-based fake news detection system	ML	Vision-and-Language BERT (VilBERT), Local Interpretable Model-Agnostic Explanations (LIME), Latent Dirichlet Allocation (LDA) topic modeling
Wang et al. (2019b)	A hybrid deep neural network model to learn the useful features from contextual information and to capture the dependencies between sequences of contextual information	DL	Recurrent and Convolutional Neural Networks (RNN and CNN)
Wang (2017)	A hybrid convolutional neural network approach for automatic fake news detection	DL	Recurrent and Convolutional Neural Networks (RNN and CNN)
Liu and Wu (2018)	An early detection approach of fake news to classify the propagation path to mine the global and local changes of user characteristics in the diffusion path	DL	Recurrent and Convolutional Neural Networks (RNN and CNN)
Mishra (2020)	Unsupervised network representation learning methods to learn user (node) embeddings from both the follower network and the retweet network and to encode the propagation path sequence	DL	RNN: (long short-term memory unit (LSTM))
Qian et al. (2018)	A Two-Level Convolutional Neural Network with User Response Generator (TCNN-URG) where TCNN captures semantic information from the article text by representing it at the sentence and word level. The URG learns a generative model of user responses to article text from historical user responses that it can use to generate responses to new articles to assist fake news detection	DL	Convolutional Neural Network (CNN)
Zhang et al. (2020)	Based on a set of explicit features extracted from the textual information, a deep diffusive network model is built to infer the credibility of news articles, creators and subjects simultaneously	DL	Deep Diffusive Network Model Learning
Goldani et al. (2021)	A capsule networks (CapsNet) approach for fake news detection using two architectures for different lengths of news statements and claims that capsule neural networks have been successful in computer vision and are receiving attention for use in Natural Language Processing (NLP)	DL	Capsule Networks (CapsNet)

Table 11 (continued)

Reference	Approach	Method	Model
Wang et al. (2019b)	An automated approach to distinguish different cases of fake news (i.e., hoaxes, irony and propaganda) while assessing and classifying news articles and claims including linguistic cues as well as user credibility and news dissemination in social media	DL, ML	Convolutional Neural Network (CNN), long Short-Term Memory (LSTM), logistic regression
Abdullah-All-Tanvir et al. (2019)	A model to recognize forged news messages from twitter posts, by figuring out how to anticipate precision appraisals, in view of computerizing forged news identification in Twitter dataset. A combination of traditional machine learning, as well as deep learning classification models, is tested to enhance the accuracy of prediction	DL, ML	Naïve Bayes, Logistic Regression, Support Vector Machine, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM)
Kaliyar et al. (2020)	An approach named (FNDNet) based on the combination between unsupervised learning algorithm GloVe and deep convolutional neural network for fake news detection	DL, ML	Deep Convolutional Neural Network (CNN), Global Vectors (GloVe)
Zhang et al. (2019a)	A hybrid approach to encode auxiliary information coming from people's replies alone in temporal order. Such auxiliary information is then used to update a priori belief generating a posteriori belief	DL, ML	Deep Learning Model, Long Short-Term Memory Neural Network (LSTM)
Deepak and Chitturi (2020)	A system that consists of live data mining in addition to the deep learning model	DL, ML	Feedforward Neural Network (FNN) and LSTM Word Vector Model
Shu et al. (2018a)	A multidimensional fake news data repository "FakeNewsNet" and conduct an exploratory analysis of the datasets to evaluate them	DL, ML	Convolutional Neural Network (CNN), Support Vector Machines (SVMs), Logistic Regression (LR), Naïve Bayes (NB)
Vereshchaka et al. (2020)	A sociocultural textual analysis, computational linguistics analysis, and textual classification using NLP, as well as deep learning models to distinguish fake from real news to mitigate the problem of disinformation	DL, NLP	Short-Term Memory (LSTM), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU)
Kapusta et al. (2019)	A sentiment and frequency analysis using both machine learning and NLP in what is called text mining to processing news content sentiment analysis and frequency analysis to compare basic text characteristics of fake and real news articles	ML, NLP	The Natural Language Toolkit (NLTK), TF-IDF
Ozbay and Alatas (2020)	A hybrid approach based on text analysis and supervised artificial intelligence for fake news detection	ML, NLP	Supervised algorithms: BayesNet, JRip, OneR, Decision Stump, ZeroR, Stochastic Gradient Descent (SGD), CV Parameter Selection (CVPS), Randomizable Filtered Classifier (RFC), Logistic Model Tree (LMT). NLP: TF weighting
Ahmed et al. (2020)	A machine learning and NLP text-based processing to identify fake news. Various features of the text are extracted through text processing and after that those features are incorporated into classification	ML, NLP	Machine learning classifiers (i.e., Passive-aggressive, Naïve Bayes and Support Vector Machine)

Table 11 (continued)

Reference	Approach	Method	Model
Abdullah-All-Tanvir et al. (2020)	A hybrid neural network approach to identify authentic news on popular Twitter threads which would outperform the traditional neural network architecture's performance. Three traditional supervised algorithms and two Deep Neural are combined to train the defined model. Some NLP concepts were also used to implement some of the traditional supervised machine learning algorithms over their dataset	ML, DL, NLP	Traditional supervised algorithm (i.e., Logistic Regression, Bayesian Classifier and Support Vector Machine). Deep Neural Networks (i.e., Recurrent Neural Network, Long Short-Term Memory LSTM). NLP concepts such as Count vectorizer and TF-IDF Vectorizer
Kaur et al. (2020)	A hybrid method to identify news articles as fake or real through finding out which classification model identifies false features accurately	ML, DL, NLP	Neural Networks (NN) and Ensemble Models. Supervised Machine Learning Classifiers such as Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Linear Models. Term Frequency-Inverse Document Frequency (TF-IDF), Count-Vectorizer (CV), Hashing-Vectorizer (HV)
Kaliyar (2018)	A fake news detection approach to classify the news article or other documents into certain or not. Natural language processing, machine learning and deep learning techniques are used to implement the defined models and to predict the accuracy of different models and classifiers	ML, DL, NLP	Machine Learning Models: Naïve Bayes, K-nearest Neighbors, Decision Tree, Random Forest. Deep Learning Networks: Shallow Convolutional Neural Networks (CNN), Very Deep Convolutional Neural Network (VDCNN), Long Short-Term Memory Network (LSTM), Gated Recurrent Unit Network (GRU). A combination of Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) and Convolutional Neural Network with Gated Recurrent Unit (CNN-LSTM)
Mahabub (2020)	An intelligent detection system to manage the classification of news as being either real or fake	ML, DL, NLP	Machine Learning: Naïve Bayes, KNN, SVM, Random Forest, Artificial Neural Network, Logistic Regression, Gradient Boosting, AdaBoost
Bahad et al. (2019)	A method based on Bi-directional LSTM-recurrent neural network to analyze the relationship between the news article headline and article body	ML, DL, NLP	Unsupervised Learning algorithm: Global Vectors (GloVe). Bi-directional LSTM-recurrent Neural Network

Author Contributions The order of authors is alphabetic as is customary in the third author's field. The lead author was Sabrine Amri, who collected and analyzed the data and wrote a first draft of the paper, all along under the supervision and tight guidance of Esma Aïmeur. Gilles Brassard reviewed, criticized and polished the work into its final form.

Funding This work is supported in part by Canada's Natural Sciences and Engineering Research Council.

Availability of data and material All the data and material are available in the papers cited in the references.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abdullah-All-Tanvir, Mahir EM, Akhter S, Huq MR (2019) Detecting fake news using machine learning and deep learning algorithms. In: 7th international conference on smart computing and communications (ICSCC), IEEE, pp 1–5 https://doi.org/10.1109/ICSCC.2019.8843612

Abdullah-All-Tanvir, Mahir EM, Huda SMA, Barua S (2020) A hybrid approach for identifying authentic news using deep learning methods on popular Twitter threads. In: International conference on artificial intelligence and signal processing (AISP), IEEE, pp 1–6 https://doi.org/10.1109/AISP48273.2020.9073583

Abu Arqoub O, Abdulateef Elega A, Efe Özad B, Dwikat H, Adedamola Oloyede F (2022) Mapping the scholarship of fake news research: a systematic review. J Pract 16(1):56–86. https://doi.org/10.1080/17512786.2020.1805791

Ahmed S, Hinkelmann K, Corradini F (2020) Development of fake news model using machine learning through natural language processing. Int J Comput Inf Eng 14(12):454–460

Aïmeur E, Brassard G, Rioux J (2013) Data privacy: an end-user perspective. Int J Comput Netw Commun Secur 1(6):237–250

Aïmeur E, Hage H, Amri S (2018) The scourge of online deception in social networks. In: 2018 international conference on computational science and computational intelligence (CSCI), IEEE, pp 1266–1271 https://doi.org/10.1109/CSCI46756.2018.00244

Alemanno A (2018) How to counter fake news? A taxonomy of anti-fake news approaches. Eur J Risk Regul 9(1):1–5. https://doi.org/10.1017/err.2018.12

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–36. https://doi.org/10.1257/jep.31.2.211

Allen J, Howland B, Mobius M, Rothschild D, Watts DJ (2020) Evaluating the fake news problem at the scale of the information ecosystem. Sci Adv. https://doi.org/10.1126/sciadv.aay3539

Allington D, Duffy B, Wessely S, Dhavan N, Rubin J (2020) Health-protective behaviour, social media usage and conspiracy belief during the Covid-19 public health emergency. Psychol Med. https://doi.org/10.1017/S003329172000224X

Alonso-Galbán P, Alemañy-Castilla C (2022) Curbing misinformation and disinformation in the Covid-19 era: a view from cuba. MEDICC Rev 22:45–46 https://doi.org/10.37757/MR2020.V22.N2.12

Altay S, Hacquin AS, Mercier H (2022) Why do so few people share fake news? It hurts their reputation. New Media Soc 24(6):1303–1324. https://doi.org/10.1177/1461444820969893

Amri S, Sallami D, Aïmeur E (2022) Exmulf: an explainable multimodal content-based fake news detection system. In: International symposium on foundations and practice of security. Springer, Berlin, pp 177–187. https://doi.org/10.1109/IJCNN48605.2020.9206973

Andersen J, Søe SO (2020) Communicative actions we live by: the problem with fact-checking, tagging or flagging fake news-the case of Facebook. Eur J Commun 35(2):126–139. https://doi.org/10.1177/0267323119894489

Apuke OD, Omar B (2021) Fake news and Covid-19: modelling the predictors of fake news sharing among social media users. Telematics Inform 56:101475. https://doi.org/10.1016/j.tele.2020.101475

Apuke OD, Omar B, Tunca EA, Gever CV (2022) The effect of visual multimedia instructions against fake news spread: a quasi-experimental study with Nigerian students. J Librariansh Inf Sci. https://doi.org/10.1177/09610006221096477

Aswani R, Ghrera S, Kar AK, Chandra S (2017) Identifying buzz in social media: a hybrid approach using artificial bee colony and k-nearest neighbors for outlier detection. Soc Netw Anal Min 7(1):1–10. https://doi.org/10.1007/s13278-017-0461-2

Avram M, Micallef N, Patil S, Menczer F (2020) Exposure to social engagement metrics increases vulnerability to misinformation. arXiv preprint arxiv:2005.04682, https://doi.org/10.37016/mr-2020-033

Badawy A, Lerman K, Ferrara E (2019) Who falls for online political manipulation? In: Companion proceedings of the 2019 world wide web conference, pp 162–168 https://doi.org/10.1145/3308560.3316494

Bahad P, Saxena P, Kamal R (2019) Fake news detection using bi-directional LSTM-recurrent neural network. Procedia Comput Sci 165:74–82. https://doi.org/10.1016/j.procs.2020.01.072

Bakdash J, Sample C, Rankin M, Kantarcioglu M, Holmes J, Kase S, Zaroukian E, Szymanski B (2018) The future of deception: machine-generated and manipulated images, video, and audio? In: 2018 international workshop on social sensing (SocialSens), IEEE, pp 2–2 https://doi.org/10.1109/SocialSens.2018.00009

Balmas M (2014) When fake news becomes real: combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. Commun Res 41(3):430–454. https://doi.org/10.1177/0093650212453600

Baptista JP, Gradim A (2020) Understanding fake news consumption: a review. Soc Sci. https://doi.org/10.3390/socsci9100185

Baptista JP, Gradim A (2022) A working definition of fake news. Encyclopedia 2(1):632–645. https://doi.org/10.3390/encyclopedia2010043

Bastick Z (2021) Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. Comput Hum Behav 116:106633. https://doi.org/10.1016/j.chb.2020.106633

Batailler C, Brannon SM, Teas PE, Gawronski B (2022) A signal detection approach to understanding the identification of fake news. Perspect Psychol Sci 17(1):78–98. https://doi.org/10.1177/1745691620986135

Bessi A, Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. First Monday 21(11-7). https://doi.org/10.5210/fm.v21i11.7090

Bhattacharjee A, Shu K, Gao M, Liu H (2020) Disinformation in the online information ecosystem: detection, mitigation and challenges. arXiv preprint arXiv:2010.09113

Bhuiyan MM, Zhang AX, Sehat CM, Mitra T (2020) Investigating differences in crowdsourced news credibility assessment: raters,

tasks, and expert criteria. Proc ACM Hum Comput Interact 4(CSCW2):1–26. https://doi.org/10.1145/3415164
Bode L, Vraga EK (2015) In related news, that was wrong: the correction of misinformation through related stories functionality in social media. J Commun 65(4):619–638. https://doi.org/10.1111/jcom.12166
Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55. https://doi.org/10.1016/j.ins.2019.05.035
Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 10(1):1–14. https://doi.org/10.1038/s41467-018-07761-2
Brashier NM, Pennycook G, Berinsky AJ, Rand DG (2021) Timing matters when correcting fake news. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.2020043118
Brewer PR, Young DG, Morreale M (2013) The impact of real news about "fake news'': intertextual processes and political satire. Int J Public Opin Res 25(3):323–343. https://doi.org/10.1093/ijpor/edt015
Bringula RP, Catacutan-Bangit AE, Garcia MB, Gonzales JPS, Valderama AMC (2022) "Who is gullible to political disinformation?'' Predicting susceptibility of university students to fake news. J Inf Technol Polit 19(2):165–179. https://doi.org/10.1080/19331681.2021.1945988
Buccafurri F, Lax G, Nicolazzo S, Nocera A (2017) Tweetchain: an alternative to blockchain for crowd-based applications. In: International conference on web engineering, Springer, Berlin, pp 386–393. https://doi.org/10.1007/978-3-319-60131-1_24
Burshtein S (2017) The true story on fake news. Intell Prop J 29(3):397–446
Cardaioli M, Cecconello S, Conti M, Pajola L, Turrin F (2020) Fake news spreaders profiling through behavioural analysis. In: CLEF (working notes)
Cardoso Durier da Silva F, Vieira R, Garcia AC (2019) Can machines learn to detect fake news? A survey focused on social media. In: Proceedings of the 52nd Hawaii international conference on system sciences. https://doi.org/10.24251/HICSS.2019.332
Carmi E, Yates SJ, Lockley E, Pawluczuk A (2020) Data citizenship: rethinking data literacy in the age of disinformation, misinformation, and malinformation. Intern Policy Rev 9(2):1–22 https://doi.org/10.14763/2020.2.1481
Celliers M, Hattingh M (2020) A systematic review on fake news themes reported in literature. In: Conference on e-Business, e-Services and e-Society. Springer, Berlin, pp 223–234. https://doi.org/10.1007/978-3-030-45002-1_19
Chen Y, Li Q, Wang H (2018) Towards trusted social networks with blockchain technology. arXiv preprint arXiv:1801.02796
Cheng L, Guo R, Shu K, Liu H (2020) Towards causal understanding of fake news dissemination. arXiv preprint arXiv:2010.10580
Chiu MM, Oh YW (2021) How fake news differs from personal lies. Am Behav Sci 65(2):243–258. https://doi.org/10.1177/0002764220910243
Chung M, Kim N (2021) When I learn the news is false: how fact-checking information stems the spread of fake news via third-person perception. Hum Commun Res 47(1):1–24. https://doi.org/10.1093/hcr/hqaa010
Clarke J, Chen H, Du D, Hu YJ (2020) Fake news, investor attention, and market reaction. Inf Syst Res. https://doi.org/10.1287/isre.2019.0910
Clayton K, Blair S, Busam JA, Forstner S, Glance J, Green G, Kawata A, Kovvuri A, Martin J, Morgan E et al (2020) Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Polit Behav 42(4):1073–1095. https://doi.org/10.1007/s11109-019-09533-0
Collins B, Hoang DT, Nguyen NT, Hwang D (2020) Fake news types and detection models on social media a state-of-the-art survey. In: Asian conference on intelligent information and database systems. Springer, Berlin, pp 562–573 https://doi.org/10.1007/978-981-15-3380-8_49
Conroy NK, Rubin VL, Chen Y (2015) Automatic deception detection: methods for finding fake news. Proc Assoc Inf Sci Technol 52(1):1–4. https://doi.org/10.1002/pra2.2015.145052010082
Cooke NA (2017) Posttruth, truthiness, and alternative facts: Information behavior and critical information consumption for a new age. Libr Q 87(3):211–221. https://doi.org/10.1086/692298
Coscia M, Rossi L (2020) Distortions of political bias in crowdsourced misinformation flagging. J R Soc Interface 17(167):20200020. https://doi.org/10.1098/rsif.2020.0020
Dame Adjin-Tettey T (2022) Combating fake news, disinformation, and misinformation: experimental evidence for media literacy education. Cogent Arts Human 9(1):2037229. https://doi.org/10.1080/23311983.2022.2037229
Deepak S, Chitturi B (2020) Deep neural approach to fake-news identification. Procedia Comput Sci 167:2236–2243. https://doi.org/10.1016/j.procs.2020.03.276
de Cock Buning M (2018) A multi-dimensional approach to disinformation: report of the independent high level group on fake news and online disinformation. Publications Office of the European Union
Del Vicario M, Quattrociocchi W, Scala A, Zollo F (2019) Polarization and fake news: early warning of potential misinformation targets. ACM Trans Web (TWEB) 13(2):1–22. https://doi.org/10.1145/3316809
Demuyakor J, Opata EM (2022) Fake news on social media: predicting which media format influences fake news most on facebook. J Intell Commun. https://doi.org/10.54963/jic.v2i1.56
Derakhshan H, Wardle C (2017) Information disorder: definitions. In: Understanding and addressing the disinformation ecosystem, pp 5–12
Desai AN, Ruidera D, Steinbrink JM, Granwehr B, Lee DH (2022) Misinformation and disinformation: the potential disadvantages of social media in infectious disease and how to combat them. Clin Infect Dis 74(Supplement–3):e34–e39. https://doi.org/10.1093/cid/ciac109
Di Domenico G, Sit J, Ishizaka A, Nunan D (2021) Fake news, social media and marketing: a systematic review. J Bus Res 124:329–341. https://doi.org/10.1016/j.jbusres.2020.11.037
Dias N, Pennycook G, Rand DG (2020) Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. Harv Kennedy School Misinform Rev. https://doi.org/10.37016/mr-2020-001
DiCicco KW, Agarwal N (2020) Blockchain technology-based solutions to fight misinformation: a survey. In: Disinformation, misinformation, and fake news in social media. Springer, Berlin, pp 267–281, https://doi.org/10.1007/978-3-030-42699-6_14
Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, Deravi F (2019) Understanding conspiracy theories. Polit Psychol 40:3–35. https://doi.org/10.1111/pops.12568
Edgerly S, Mourão RR, Thorson E, Tham SM (2020) When do audiences verify? How perceptions about message and source influence audience verification of news headlines. J Mass Commun Q 97(1):52–71. https://doi.org/10.1177/1077699019864680
Egelhofer JL, Lecheler S (2019) Fake news as a two-dimensional phenomenon: a framework and research agenda. Ann Int Commun Assoc 43(2):97–116. https://doi.org/10.1080/23808985.2019.1602782
Elhadad MK, Li KF, Gebali F (2019) A novel approach for selecting hybrid features from online news textual metadata for fake news detection. In: International conference on p2p, parallel, grid,

cloud and internet computing. Springer, Berlin, pp 914–925, https://doi.org/10.1007/978-3-030-33509-0_86
ERGA (2018) Fake news, and the information disorder. European Broadcasting Union (EBU)
ERGA (2021) Notions of disinformation and related concepts. European Regulators Group for Audiovisual Media Services (ERGA)
Escolà-Gascón Á (2021) New techniques to measure lie detection using Covid-19 fake news and the Multivariable Multiaxial Suggestibility Inventory-2 (MMSI-2). Comput Hum Behav Rep 3:100049. https://doi.org/10.1016/j.chbr.2020.100049
Fazio L (2020) Pausing to consider why a headline is true or false can help reduce the sharing of false news. Harv Kennedy School Misinformation Rev. https://doi.org/10.37016/mr-2020-009
Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104. https://doi.org/10.1145/2818717
Flynn D, Nyhan B, Reifler J (2017) The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. Polit Psychol 38:127–150. https://doi.org/10.1111/pops.12394
Fraga-Lamas P, Fernández-Caramés TM (2020) Fake news, disinformation, and deepfakes: leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality. IT Prof 22(2):53–59. https://doi.org/10.1109/MITP.2020.2977589
Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, Jenner L, Teale AL, Carr L, Mulhall S et al (2020) Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. Psychol Med. https://doi.org/10.1017/S0033291720001890
Friggeri A, Adamic L, Eckles D, Cheng J (2014) Rumor cascades. In: Proceedings of the international AAAI conference on web and social media
García SA, García GG, Prieto MS, Moreno Guerrero AJ, Rodríguez Jiménez C (2020) The impact of term fake news on the scientific community. Scientific performance and mapping in web of science. Soc Sci. https://doi.org/10.3390/socsci9050073
Garrett RK, Bond RM (2021) Conservatives' susceptibility to political misperceptions. Sci Adv. https://doi.org/10.1126/sciadv.abf1234
Giachanou A, Ríssola EA, Ghanem B, Crestani F, Rosso P (2020) The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: International conference on applications of natural language to information systems. Springer, Berlin, pp 181–192 https://doi.org/10.1007/978-3-030-51310-8_17
Golbeck J, Mauriello M, Auxier B, Bhanushali KH, Bonk C, Bouzaghrane MA, Buntain C, Chanduka R, Cheakalos P, Everett JB et al (2018) Fake news vs satire: a dataset and analysis. In: Proceedings of the 10th ACM conference on web science, pp 17–21, https://doi.org/10.1145/3201064.3201100
Goldani MH, Momtazi S, Safabakhsh R (2021) Detecting fake news with capsule neural networks. Appl Soft Comput 101:106991. https://doi.org/10.1016/j.asoc.2020.106991
Goldstein I, Yang L (2019) Good disclosure, bad disclosure. J Financ Econ 131(1):118–138. https://doi.org/10.1016/j.jfineco.2018.08.004
Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 US presidential election. Science 363(6425):374–378. https://doi.org/10.1126/science.aau2706
Guadagno RE, Guttieri K (2021) Fake news and information warfare: an examination of the political and psychological processes from the digital sphere to the real world. In: Research anthology on fake news, political warfare, and combatting the spread of misinformation. IGI Global, pp 218–242 https://doi.org/10.4018/978-1-7998-7291-7.ch013
Guess A, Nagler J, Tucker J (2019) Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv. https://doi.org/10.1126/sciadv.aau4586
Guo C, Cao J, Zhang X, Shu K, Yu M (2019) Exploiting emotions for fake news detection on social media. arXiv preprint arXiv:1903.01728
Guo B, Ding Y, Yao L, Liang Y, Yu Z (2020) The future of false information detection on social media: new perspectives and trends. ACM Comput Surv (CSUR) 53(4):1–36. https://doi.org/10.1145/3393880
Gupta A, Li H, Farnoush A, Jiang W (2022) Understanding patterns of covid infodemic: a systematic and pragmatic approach to curb fake news. J Bus Res 140:670–683. https://doi.org/10.1016/j.jbusres.2021.11.032
Ha L, Andreu Perez L, Ray R (2021) Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: disciplinary contribution, topics, and impact. Am Behav Sci 65(2):290–315. https://doi.org/10.1177/0002764219869402
Habib A, Asghar MZ, Khan A, Habib A, Khan A (2019) False information detection in online content and its role in decision making: a systematic literature review. Soc Netw Anal Min 9(1):1–20. https://doi.org/10.1007/s13278-019-0595-5
Hage H, Aïmeur E, Guedidi A (2021) Understanding the landscape of online deception. In: Research anthology on fake news, political warfare, and combatting the spread of misinformation. IGI Global, pp 39–66. https://doi.org/10.4018/978-1-7998-2543-2.ch014
Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ (2021) An ensemble machine learning approach through effective feature extraction to classify fake news. Futur Gener Comput Syst 117:47–58. https://doi.org/10.1016/j.future.2020.11.022
Hamdi T, Slimi H, Bounhas I, Slimani Y (2020) A hybrid approach for fake news detection in Twitter based on user features and graph embedding. In: International conference on distributed computing and internet technology. Springer, Berlin, pp 266–280. https://doi.org/10.1007/978-3-030-36987-3_17
Hameleers M (2022) Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the us and netherlands. Inf Commun Soc 25(1):110–126. https://doi.org/10.1080/1369118X.2020.1764603
Hameleers M, Powell TE, Van Der Meer TG, Bos L (2020) A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. Polit Commun 37(2):281–301. https://doi.org/10.1080/10584609.2019.1674979
Hameleers M, Brosius A, de Vreese CH (2022) Whom to trust? media exposure patterns of citizens with perceptions of misinformation and disinformation related to the news media. Eur J Commun. https://doi.org/10.1177/02673231211072667
Hartley K, Vu MK (2020) Fighting fake news in the Covid-19 era: policy insights from an equilibrium model. Policy Sci 53(4):735–758. https://doi.org/10.1007/s11077-020-09405-z
Hasan HR, Salah K (2019) Combating deepfake videos using blockchain and smart contracts. IEEE Access 7:41596–41606. https://doi.org/10.1109/ACCESS.2019.2905689
Hiriyannaiah S, Srinivas A, Shetty GK, Siddesh G, Srinivasa K (2020) A computationally intelligent agent for detecting fake news using generative adversarial networks. Hybrid computational intelligence: challenges and applications. pp 69–96 https://doi.org/10.1016/B978-0-12-818699-2.00004-4
Hosseinimotlagh S, Papalexakis EE (2018) Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In: Proceedings of the workshop on misinformation and misbehavior mining on the web (MIS2)

Huckle S, White M (2017) Fake news: a technological approach to proving the origins of content, using blockchains. Big Data 5(4):356–371. https://doi.org/10.1089/big.2017.0071

Huffaker JS, Kummerfeld JK, Lasecki WS, Ackerman MS (2020) Crowdsourced detection of emotionally manipulative language. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp 1–14 https://doi.org/10.1145/3313831.3376375

Ireton C, Posetti J (2018) Journalism, fake news & disinformation: handbook for journalism education and training. UNESCO Publishing, Paris

Islam MR, Liu S, Wang X, Xu G (2020) Deep learning for misinformation detection on online social networks: a survey and new perspectives. Soc Netw Anal Min 10(1):1–20. https://doi.org/10.1007/s13278-020-00696-x

Ismailov M, Tsikerdekis M, Zeadally S (2020) Vulnerabilities to online social network identity deception detection research and recommendations for mitigation. Fut Internet 12(9):148. https://doi.org/10.3390/fi12090148

Jakesch M, Koren M, Evtushenko A, Naaman M (2019) The role of source and expressive responding in political news evaluation. In: Computation and journalism symposium

Jamieson KH (2020) Cyberwar: how Russian hackers and trolls helped elect a president: what we don't, can't, and do know. Oxford University Press, Oxford. https://doi.org/10.1093/poq/nfy049

Jiang S, Chen X, Zhang L, Chen S, Liu H (2019) User-characteristic enhanced model for fake news detection in social media. In: CCF International conference on natural language processing and Chinese computing, Springer, Berlin, pp 634–646. https://doi.org/10.1007/978-3-030-32233-5_49

Jin Z, Cao J, Zhang Y, Luo J (2016) News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the AAAI conference on artificial intelligence

Jing TW, Murugesan RK (2018) A theoretical framework to build trust and prevent fake news in social media using blockchain. In: International conference of reliable information and communication technology. Springer, Berlin, pp 955–962, https://doi.org/10.1007/978-3-319-99007-1_88

Jones-Jang SM, Mortensen T, Liu J (2021) Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. Am Behav Sci 65(2):371–388. https://doi.org/10.1177/0002764219869406

Jungherr A, Schroeder R (2021) Disinformation and the structural transformations of the public arena: addressing the actual challenges to democracy. Soc Media Soc. https://doi.org/10.1177/2056305121988928

Kaliyar RK (2018) Fake news detection using a deep neural network. In: 2018 4th international conference on computing communication and automation (ICCCA), IEEE, pp 1–7 https://doi.org/10.1109/CCAA.2018.8777343

Kaliyar RK, Goswami A, Narang P, Sinha S (2020) Fndnet—a deep convolutional neural network for fake news detection. Cogn Syst Res 61:32–44. https://doi.org/10.1016/j.cogsys.2019.12.005

Kapantai E, Christopoulou A, Berberidis C, Peristeras V (2021) A systematic literature review on disinformation: toward a unified taxonomical framework. New Media Soc 23(5):1301–1326. https://doi.org/10.1177/1461444820959296

Kapusta J, Benko L, Munk M (2019) Fake news identification based on sentiment and frequency analysis. In: International conference Europe middle east and North Africa information systems and technologies to support learning. Springer, Berlin, pp 400–409, 10.1007/978-3-030-36778-7_44

Kaur S, Kumar P, Kumaraguru P (2020) Automating fake news detection system using multi-level voting model. Soft Comput 24(12):9049–9069. https://doi.org/10.1007/s00500-019-04436-y

Khan SA, Alkawaz MH, Zangana HM (2019) The use and abuse of social media for spreading fake news. In: 2019 IEEE international conference on automatic control and intelligent systems (I2CACIS), IEEE, pp 145–148. https://doi.org/10.1109/I2CACIS.2019.8825029

Kim J, Tabibian B, Oh A, Schölkopf B, Gomez-Rodriguez M (2018) Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 324–332. https://doi.org/10.1145/3159652.3159734

Klein D, Wueller J (2017) Fake news: a legal perspective. J Internet Law 20(10):5–13

Kogan S, Moskowitz TJ, Niessner M (2019) Fake news: evidence from financial markets. Available at SSRN 3237763

Kuklinski JH, Quirk PJ, Jerit J, Schwieder D, Rich RF (2000) Misinformation and the currency of democratic citizenship. J Polit 62(3):790–816. https://doi.org/10.1111/0022-3816.00033

Kumar S, Shah N (2018) False information on web and social media: a survey. arXiv preprint arXiv:1804.08559

Kumar S, West R, Leskovec J (2016) Disinformation on the web: impact, characteristics, and detection of Wikipedia hoaxes. In: Proceedings of the 25th international conference on world wide web, pp 591–602. https://doi.org/10.1145/2872427.2883085

La Barbera D, Roitero K, Demartini G, Mizzaro S, Spina D (2020) Crowdsourcing truthfulness: the impact of judgment scale and assessor bias. In: European conference on information retrieval. Springer, Berlin, pp 207–214. https://doi.org/10.1007/978-3-030-45442-5_26

Lanius C, Weber R, MacKenzie WI (2021) Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey. Soc Netw Anal Min 11(1):1–15. https://doi.org/10.1007/s13278-021-00739-x

Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D et al (2018) The science of fake news. Science 359(6380):1094–1096. https://doi.org/10.1126/science.aao2998

Le T, Shu K, Molina MD, Lee D, Sundar SS, Liu H (2019) 5 sources of clickbaits you should know! Using synthetic clickbaits to improve prediction and distinguish between bot-generated and human-written headlines. In: 2019 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 33–40. https://doi.org/10.1145/3341161.3342875

Lewandowsky S (2020) Climate change, disinformation, and how to combat it. In: Annual Review of Public Health 42. https://doi.org/10.1146/annurev-publhealth-090419-102409

Liu Y, Wu YF (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence, pp 354–361

Luo M, Hancock JT, Markowitz DM (2022) Credibility perceptions and detection accuracy of fake news headlines on social media: effects of truth-bias and endorsement cues. Commun Res 49(2):171–195. https://doi.org/10.1177/0093650220921321

Lutzke L, Drummond C, Slovic P, Árvai J (2019) Priming critical thinking: simple interventions limit the influence of fake news about climate change on Facebook. Glob Environ Chang 58:101964. https://doi.org/10.1016/j.gloenvcha.2019.101964

Maertens R, Anseel F, van der Linden S (2020) Combatting climate change misinformation: evidence for longevity of inoculation and consensus messaging effects. J Environ Psychol 70:101455. https://doi.org/10.1016/j.jenvp.2020.101455

Mahabub A (2020) A robust technique of fake news detection using ensemble voting classifier and comparison with other classifiers. SN Applied Sciences 2(4):1–9. https://doi.org/10.1007/s42452-020-2326-y

Mahbub S, Pardede E, Kayes A, Rahayu W (2019) Controlling astroturfing on the internet: a survey on detection techniques and research challenges. Int J Web Grid Serv 15(2):139–158. https://doi.org/10.1504/IJWGS.2019.099561

Marsden C, Meyer T, Brown I (2020) Platform values and democratic elections: how can the law regulate digital disinformation? Comput Law Secur Rev 36:105373. https://doi.org/10.1016/j.clsr.2019.105373

Masciari E, Moscato V, Picariello A, Sperlí G (2020) Detecting fake news by image analysis. In: Proceedings of the 24th symposium on international database engineering and applications, pp 1–5. https://doi.org/10.1145/3410566.3410599

Mazzeo V, Rapisarda A (2022) Investigating fake and reliable news sources using complex networks analysis. Front Phys 10:886544. https://doi.org/10.3389/fphy.2022.886544

McGrew S (2020) Learning to evaluate: an intervention in civic online reasoning. Comput Educ 145:103711. https://doi.org/10.1016/j.compedu.2019.103711

McGrew S, Breakstone J, Ortega T, Smith M, Wineburg S (2018) Can students evaluate online sources? Learning from assessments of civic online reasoning. Theory Res Soc Educ 46(2):165–193. https://doi.org/10.1080/00933104.2017.1416320

Meel P, Vishwakarma DK (2020) Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl 153:112986. https://doi.org/10.1016/j.eswa.2019.112986

Meese J, Frith J, Wilken R (2020) Covid-19, 5G conspiracies and infrastructural futures. Media Int Aust 177(1):30–46. https://doi.org/10.1177/1329878X20952165

Metzger MJ, Hartsell EH, Flanagin AJ (2020) Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. Commun Res 47(1):3–28. https://doi.org/10.1177/0093650215613136

Micallef N, He B, Kumar S, Ahamad M, Memon N (2020) The role of the crowd in countering misinformation: a case study of the Covid-19 infodemic. arXiv preprint arXiv:2011.05773

Mihailidis P, Viotty S (2017) Spreadable spectacle in digital culture: civic expression, fake news, and the role of media literacies in "post-fact society. Am Behav Sci 61(4):441–454. https://doi.org/10.1177/0002764217701217

Mishra R (2020) Fake news detection using higher-order user to user mutual-attention progression in propagation paths. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 652–653

Mishra S, Shukla P, Agarwal R (2022) Analyzing machine learning enabled fake news detection techniques for diversified datasets. Wirel Commun Mobile Comput. https://doi.org/10.1155/2022/1575365

Molina MD, Sundar SS, Le T, Lee D (2021) "Fake news'' is not simply false information: a concept explication and taxonomy of online content. Am Behav Sci 65(2):180–212. https://doi.org/10.1177/0002764219878224

Moro C, Birt JR (2022) Review bombing is a dirty practice, but research shows games do benefit from online feedback. Conversation. https://research.bond.edu.au/en/publications/review-bombing-is-a-dirty-practice-but-research-shows-games-do-be

Mustafaraj E, Metaxas PT (2017) The fake news spreading plague: was it preventable? In: Proceedings of the 2017 ACM on web science conference, pp 235–239. https://doi.org/10.1145/3091478.3091523

Nagel TW (2022) Measuring fake news acumen using a news media literacy instrument. J Media Liter Educ 14(1):29–42. https://doi.org/10.23860/JMLE-2022-14-1-3

Nakov P (2020) Can we spot the "fake news" before it was even written? arXiv preprint arXiv:2008.04374

Nekmat E (2020) Nudge effect of fact-check alerts: source influence and media skepticism on sharing of news misinformation in social media. Soc Media Soc. https://doi.org/10.1177/2056305119897322

Nygren T, Brounéus F, Svensson G (2019) Diversity and credibility in young people's news feeds: a foundation for teaching and learning citizenship in a digital era. J Soc Sci Educ 18(2):87–109. https://doi.org/10.4119/jsse-917

Nyhan B, Reifler J (2015) Displacing misinformation about events: an experimental test of causal corrections. J Exp Polit Sci 2(1):81–93. https://doi.org/10.1017/XPS.2014.22

Nyhan B, Porter E, Reifler J, Wood TJ (2020) Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. Polit Behav 42(3):939–960. https://doi.org/10.1007/s11109-019-09528-x

Nyow NX, Chua HN (2019) Detecting fake news with tweets' properties. In: 2019 IEEE conference on application, information and network security (AINS), IEEE, pp 24–29. https://doi.org/10.1109/AINS47559.2019.8968706

Ochoa IS, de Mello G, Silva LA, Gomes AJ, Fernandes AM, Leithardt VRQ (2019) Fakechain: a blockchain architecture to ensure trust in social media networks. In: International conference on the quality of information and communications technology. Springer, Berlin, pp 105–118. https://doi.org/10.1007/978-3-030-29238-6_8

Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A 540:123174. https://doi.org/10.1016/j.physa.2019.123174

Ozturk P, Li H, Sakamoto Y (2015) Combating rumor spread on social media: the effectiveness of refutation and warning. In: 2015 48th Hawaii international conference on system sciences, IEEE, pp 2406–2414. https://doi.org/10.1109/HICSS.2015.288

Parikh SB, Atrey PK (2018) Media-rich fake news detection: a survey. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR), IEEE, pp 436–441.https://doi.org/10.1109/MIPR.2018.00093

Parrish K (2018) Deep learning & machine learning: what's the difference? Online: https://parsers.me/deep-learning-machine-learning-whats-the-difference/. Accessed 20 May 2020

Paschen J (2019) Investigating the emotional appeal of fake news using artificial intelligence and human contributions. J Prod Brand Manag 29(2):223–233. https://doi.org/10.1108/JPBM-12-2018-2179

Pathak A, Srihari RK (2019) Breaking! Presenting fake news corpus for automated fact checking. In: Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop, pp 357–362

Peng J, Detchon S, Choo KKR, Ashman H (2017) Astroturfing detection in social media: a binary n-gram-based approach. Concurr Comput: Pract Exp 29(17):e4013. https://doi.org/10.1002/cpe.4013

Pennycook G, Rand DG (2019) Fighting misinformation on social media using crowdsourced judgments of news source quality. Proc Natl Acad Sci 116(7):2521–2526. https://doi.org/10.1073/pnas.1806781116

Pennycook G, Rand DG (2020) Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. J Pers 88(2):185–200. https://doi.org/10.1111/jopy.12476

Pennycook G, Bear A, Collins ET, Rand DG (2020a) The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. Manag Sci 66(11):4944–4957. https://doi.org/10.1287/mnsc.2019.3478

Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG (2020b) Fighting Covid-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 31(7):770–780. https://doi.org/10.1177/0956797620939054

Potthast M, Kiesel J, Reinartz K, Bevendorff J, Stein B (2017) A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638

Previti M, Rodriguez-Fernandez V, Camacho D, Carchiolo V, Malgeri M (2020) Fake news detection using time series and user features classification. In: International conference on the applications of evolutionary computation (Part of EvoStar), Springer, Berlin, pp 339–353. https://doi.org/10.1007/978-3-030-43722-0_22

Przybyla P (2020) Capturing the style of fake news. In: Proceedings of the AAAI conference on artificial intelligence, pp 490–497. https://doi.org/10.1609/aaai.v34i01.5386

Qayyum A, Qadir J, Janjua MU, Sher F (2019) Using blockchain to rein in the new post-truth world and check the spread of fake news. IT Prof 21(4):16–24. https://doi.org/10.1109/MITP.2019.2910503

Qian F, Gong C, Sharma K, Liu Y (2018) Neural user response generator: fake news detection with collective user intelligence. In: IJCAI, vol 18, pp 3834–3840. https://doi.org/10.24963/ijcai.2018/533

Raza S, Ding C (2022) Fake news detection based on news content and social contexts: a transformer-based approach. Int J Data Sci Anal 13(4):335–362. https://doi.org/10.1007/s41060-021-00302-z

Ricard J, Medeiros J (2020) Using misinformation as a political weapon: Covid-19 and Bolsonaro in Brazil. Harv Kennedy School misinformation Rev 1(3). https://misinforeview.hks.harvard.edu/article/using-misinformation-as-a-political-weapon-covid-19-and-bolsonaro-in-brazil/

Roozenbeek J, van der Linden S (2019) Fake news game confers psychological resistance against online misinformation. Palgrave Commun 5(1):1–10. https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek J, van der Linden S, Nygren T (2020a) Prebunking interventions based on the psychological theory of "inoculation'' can reduce susceptibility to misinformation across cultures. Harv Kennedy School Misinformation Rev. https://doi.org/10.37016//mr-2020-008

Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman AL, Recchia G, Van Der Bles AM, Van Der Linden S (2020b) Susceptibility to misinformation about Covid-19 around the world. R Soc Open Sci 7(10):201199. https://doi.org/10.1098/rsos.201199

Rubin VL, Conroy N, Chen Y, Cornwell S (2016) Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of the second workshop on computational approaches to deception detection, pp 7–17

Ruchansky N, Seo S, Liu Y (2017) Csi: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp 797–806. https://doi.org/10.1145/3132847.3132877

Schuyler AJ (2019) Regulating facts: a procedural framework for identifying, excluding, and deterring the intentional or knowing proliferation of fake news online. Univ Ill JL Technol Pol'y, vol 2019, pp 211–240

Shae Z, Tsai J (2019) AI blockchain platform for trusting news. In: 2019 IEEE 39th international conference on distributed computing systems (ICDCS), IEEE, pp 1610–1619. https://doi.org/10.1109/ICDCS.2019.00160

Shang W, Liu M, Lin W, Jia M (2018) Tracing the source of news based on blockchain. In: 2018 IEEE/ACIS 17th international conference on computer and information science (ICIS), IEEE, pp 377–381. https://doi.org/10.1109/ICIS.2018.8466516

Shao C, Ciampaglia GL, Flammini A, Menczer F (2016) Hoaxy: A platform for tracking online misinformation. In: Proceedings of the 25th international conference companion on world wide web, pp 745–750. https://doi.org/10.1145/2872518.2890098

Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):1–9. https://doi.org/10.1038/s41467-018-06930-7

Shao C, Hui PM, Wang L, Jiang X, Flammini A, Menczer F, Ciampaglia GL (2018) Anatomy of an online misinformation network. PLoS ONE 13(4):e0196087. https://doi.org/10.1371/journal.pone.0196087

Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: a survey on identification and mitigation techniques. ACM Trans Intell Syst Technol (TIST) 10(3):1–42. https://doi.org/10.1145/3305260

Sharma K, Seo S, Meng C, Rambhatla S, Liu Y (2020) Covid-19 on social media: analyzing misinformation in Twitter conversations. arXiv preprint arXiv:2003.12309

Shen C, Kasra M, Pan W, Bassett GA, Malloch Y, O'Brien JF (2019) Fake images: the effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. New Media Soc 21(2):438–463. https://doi.org/10.1177/1461444818799526

Sherman IN, Redmiles EM, Stokes JW (2020) Designing indicators to combat fake media. arXiv preprint arXiv:2010.00544

Shi P, Zhang Z, Choo KKR (2019) Detecting malicious social bots based on clickstream sequences. IEEE Access 7:28855–28862. https://doi.org/10.1109/ACCESS.2019.2901864

Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36. https://doi.org/10.1145/3137597.3137600

Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2018a) Fakenewsnet: a data repository with news content, social context and spatialtemporal information for studying fake news on social media. arXiv preprint arXiv:1809.01286, https://doi.org/10.1089/big.2020.0062

Shu K, Wang S, Liu H (2018b) Understanding user profiles on social media for fake news detection. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR), IEEE, pp 430–435. https://doi.org/10.1109/MIPR.2018.00092

Shu K, Wang S, Liu H (2019a) Beyond news contents: the role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 312–320. https://doi.org/10.1145/3289600.3290994

Shu K, Zhou X, Wang S, Zafarani R, Liu H (2019b) The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp 436–439. https://doi.org/10.1145/3341161.3342927

Shu K, Bhattacharjee A, Alatawi F, Nazer TH, Ding K, Karami M, Liu H (2020a) Combating disinformation in a social media age. Wiley Interdiscip Rev: Data Min Knowl Discov 10(6):e1385. https://doi.org/10.1002/widm.1385

Shu K, Mahudeswaran D, Wang S, Liu H (2020b) Hierarchical propagation networks for fake news detection: investigation and exploitation. Proc Int AAAI Conf Web Soc Media AAAI Press 14:626–637

Shu K, Wang S, Lee D, Liu H (2020c) Mining disinformation and fake news: concepts, methods, and recent advancements. In: Disinformation, misinformation, and fake news in social media. Springer, Berlin, pp 1–19 https://doi.org/10.1007/978-3-030-42699-6_1

Shu K, Zheng G, Li Y, Mukherjee S, Awadallah AH, Ruston S, Liu H (2020d) Early detection of fake news with multi-source weak social supervision. In: ECML/PKDD (3), pp 650–666

Singh VK, Ghosh I, Sonagara D (2021) Detecting fake news stories via multimodal analysis. J Am Soc Inf Sci 72(1):3–17. https://doi.org/10.1002/asi.24359

Sintos S, Agarwal PK, Yang J (2019) Selecting data to clean for fact checking: minimizing uncertainty vs. maximizing surprise. Proc VLDB Endowm 12(13), 2408–2421. https://doi.org/10.14778/3358701.3358708

Snow J (2017) Can AI win the war against fake news? MIT Technology Review Online: https://www.technologyreview.com/s/609717/can-ai-win-the-war-against-fake-news/. Accessed 3 Oct. 2020

Song G, Kim S, Hwang H, Lee K (2019) Blockchain-based notarization for social media. In: 2019 IEEE international conference on consumer electronics (ICCE), IEEE, pp 1–2 https://doi.org/10.1109/ICCE.2019.8661978

Starbird K, Arif A, Wilson T (2019) Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. In: Proceedings of the ACM on human–computer interaction, vol 3(CSCW), pp 1–26 https://doi.org/10.1145/3359229

Sterret D, Malato D, Benz J, Kantor L, Tompson T, Rosenstiel T, Sonderman J, Loker K, Swanson E (2018) Who shared it? How Americans decide what news to trust on social media. Technical report, Norc Working Paper Series, WP-2018-001, pp 1–24

Sutton RM, Douglas KM (2020) Conspiracy theories and the conspiracy mindset: implications for political ideology. Curr Opin Behav Sci 34:118–122. https://doi.org/10.1016/j.cobeha.2020.02.015

Tandoc EC Jr, Thomas RJ, Bishop L (2021) What is (fake) news? Analyzing news values (and more) in fake stories. Media Commun 9(1):110–119. https://doi.org/10.17645/mac.v9i1.3331

Tchakounté F, Faissal A, Atemkeng M, Ntyam A (2020) A reliable weighting scheme for the aggregation of crowd intelligence to detect fake news. Information 11(6):319. https://doi.org/10.3390/info11060319

Tchechmedjiev A, Fafalios P, Boland K, Gasquet M, Zloch M, Zapilko B, Dietze S, Todorov K (2019) Claimskg: a knowledge graph of fact-checked claims. In: International semantic web conference. Springer, Berlin, pp 309–324 https://doi.org/10.1007/978-3-030-30796-7_20

Treen KMd, Williams HT, O'Neill SJ (2020) Online misinformation about climate change. Wiley Interdiscip Rev Clim Change 11(5):e665. https://doi.org/10.1002/wcc.665

Tsang SJ (2020) Motivated fake news perception: the impact of news sources and policy support on audiences' assessment of news fakeness. J Mass Commun Q. https://doi.org/10.1177/1077699020952129

Tschiatschek S, Singla A, Gomez Rodriguez M, Merchant A, Krause A (2018) Fake news detection in social networks via crowd signals. In: Companion proceedings of the the web conference 2018, pp 517–524. https://doi.org/10.1145/3184558.3188722

Uppada SK, Manasa K, Vidhathri B, Harini R, Sivaselvan B (2022) Novel approaches to fake news and fake account detection in OSNS: user social engagement and visual content centric model. Soc Netw Anal Min 12(1):1–19. https://doi.org/10.1007/s13278-022-00878-9

Van der Linden S, Roozenbeek J (2020) Psychological inoculation against fake news. In: Accepting, sharing, and correcting misinformation, the psychology of fake news. https://doi.org/10.4324/9780429295379-11

Van der Linden S, Panagopoulos C, Roozenbeek J (2020) You are fake news: political bias in perceptions of fake news. Media Cult Soc 42(3):460–470. https://doi.org/10.1177/0163443720906992

Valenzuela S, Muñiz C, Santos M (2022) Social media and belief in misinformation in mexico: a case of maximal panic, minimal effects? Int J Press Polit. https://doi.org/10.1177/19401612221088988

Vasu N, Ang B, Teo TA, Jayakumar S, Raizal M, Ahuja J (2018) Fake news: national security in the post-truth era. RSIS

Vereshchaka A, Cosimini S, Dong W (2020) Analyzing and distinguishing fake and real news to mitigate the problem of disinformation. In: Computational and mathematical organization theory, pp 1–15. https://doi.org/10.1007/s10588-020-09307-8

Verstraete M, Bambauer DE, Bambauer JR (2017) Identifying and countering fake news. Arizona legal studies discussion paper 73(17-15). https://doi.org/10.2139/ssrn.3007971

Vilmer J, Escorcia A, Guillaume M, Herrera J (2018) Information manipulation: a challenge for our democracies. In: Report by the Policy Planning Staff (CAPS) of the ministry for europe and foreign affairs, and the institute for strategic research (RSEM) of the Ministry for the Armed Forces

Vishwakarma DK, Varshney D, Yadav A (2019) Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cogn Syst Res 58:217–229. https://doi.org/10.1016/j.cogsys.2019.07.004

Vlachos A, Riedel S (2014) Fact checking: task definition and dataset construction. In: Proceedings of the ACL 2014 workshop on language technologies and computational social science, pp 18–22. https://doi.org/10.3115/v1/W14-2508

von der Weth C, Abdul A, Fan S, Kankanhalli M (2020) Helping users tackle algorithmic threats on social media: a multimedia research agenda. In: Proceedings of the 28th ACM international conference on multimedia, pp 4425–4434. https://doi.org/10.1145/3394171.3414692

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151. https://doi.org/10.1126/science.aap9559

Vraga EK, Bode L (2017) Using expert sources to correct health misinformation in social media. Sci Commun 39(5):621–645. https://doi.org/10.1177/1075547017731776

Waldman AE (2017) The marketplace of fake news. Univ Pa J Const Law 20:845

Wang WY (2017) "Liar, liar pants on fire": a new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648

Wang L, Wang Y, de Melo G, Weikum G (2019a) Understanding archetypes of fake news via fine-grained classification. Soc Netw Anal Min 9(1):1–17. https://doi.org/10.1007/s13278-019-0580-z

Wang Y, Han H, Ding Y, Wang X, Liao Q (2019b) Learning contextual features with multi-head self-attention for fake news detection. In: International conference on cognitive computing. Springer, Berlin, pp 132–142. https://doi.org/10.1007/978-3-030-23407-2_11

Wang Y, McKee M, Torbica A, Stuckler D (2019c) Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 240:112552. https://doi.org/10.1016/j.socscimed.2019.112552

Wang Y, Yang W, Ma F, Xu J, Zhong B, Deng Q, Gao J (2020) Weak supervision for fake news detection via reinforcement learning. In: Proceedings of the AAAI conference on artificial intelligence, pp 516–523. https://doi.org/10.1609/aaai.v34i01.5389

Wardle C (2017) Fake news. It's complicated. Online: https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79. Accessed 3 Oct 2020

Wardle C (2018) The need for smarter definitions and practical, timely empirical research on information disorder. Digit J 6(8):951–963. https://doi.org/10.1080/21670811.2018.1502047

Wardle C, Derakhshan H (2017) Information disorder: toward an interdisciplinary framework for research and policy making. Council Eur Rep 27:1–107

Weiss AP, Alwan A, Garcia EP, Garcia J (2020) Surveying fake news: assessing university faculty's fragmented definition of fake news and its impact on teaching critical thinking. Int J Educ Integr 16(1):1–30. https://doi.org/10.1007/s40979-019-0049-x

Wu L, Liu H (2018) Tracing fake-news footprints: characterizing social media messages by how they propagate. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 637–645. https://doi.org/10.1145/3159652.3159677

Wu L, Rao Y (2020) Adaptive interaction fusion networks for fake news detection. arXiv preprint arXiv:2004.10009

Wu L, Morstatter F, Carley KM, Liu H (2019) Misinformation in social media: definition, manipulation, and detection. ACM SIGKDD Explor Newsl 21(2):80–90. https://doi.org/10.1145/3373464.3373475

Wu Y, Ngai EW, Wu P, Wu C (2022) Fake news on the internet: a literature review, synthesis and directions for future research. Intern Res. https://doi.org/10.1108/INTR-05-2021-0294

Xu K, Wang F, Wang H, Yang B (2019) Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci Technol 25(1):20–27. https://doi.org/10.26599/TST.2018.9010139

Yang F, Pentyala SK, Mohseni S, Du M, Yuan H, Linder R, Ragan ED, Ji S, Hu X (2019a) Xfake: explainable fake news detector with visualizations. In: The world wide web conference, pp 3600–3604. https://doi.org/10.1145/3308558.3314119

Yang X, Li Y, Lyu S (2019b) Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 8261–8265. https://doi.org/10.1109/ICASSP.2019.8683164

Yaqub W, Kakhidze O, Brockman ML, Memon N, Patil S (2020) Effects of credibility indicators on social media news sharing intent. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp 1–14. https://doi.org/10.1145/3313831.3376213

Yavary A, Sajedi H, Abadeh MS (2020) Information verification in social networks based on user feedback and news agencies. Soc Netw Anal Min 10(1):1–8. https://doi.org/10.1007/s13278-019-0616-4

Yazdi KM, Yazdi AM, Khodayi S, Hou J, Zhou W, Saedy S (2020) Improving fake news detection using k-means and support vector machine approaches. Int J Electron Commun Eng 14(2):38–42. https://doi.org/10.5281/zenodo.3669287

Zannettou S, Sirivianos M, Blackburn J, Kourtellis N (2019) The web of false information: rumors, fake news, hoaxes, clickbait, and various other shenanigans. J Data Inf Qual (JDIQ) 11(3):1–37. https://doi.org/10.1145/3309699

Zellers R, Holtzman A, Rashkin H, Bisk Y, Farhadi A, Roesner F, Choi Y (2019) Defending against neural fake news. arXiv preprint arXiv:1905.12616

Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. Inf Process Manag 57(2):102025. https://doi.org/10.1016/j.ipm.2019.03.004

Zhang J, Dong B, Philip SY (2020) Fakedetector: effective fake news detection with deep diffusive neural network. In: 2020 IEEE 36th international conference on data engineering (ICDE), IEEE, pp 1826–1829. 10.1109/ICDE48307.2020.00180

Zhang Q, Lipani A, Liang S, Yilmaz E (2019a) Reply-aided detection of misinformation via Bayesian deep learning. In: The world wide web conference, pp 2333–2343. https://doi.org/10.1145/3308558.3313718

Zhang X, Karaman S, Chang SF (2019b) Detecting and simulating artifacts in GAN fake images. In: 2019 IEEE international workshop on information forensics and security (WIFS), IEEE, pp 1–6 https://doi.org/10.1109/WIFS47025.2019.9035107

Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput Surv (CSUR) 53(5):1–40. https://doi.org/10.1145/3395046

Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and resolution of rumours in social media: a survey. ACM Comput Surv (CSUR) 51(2):1–36. https://doi.org/10.1145/3161603