A Literature Review on Detecting, Verifying, and Mitigating Online Misinformation

Arezo Bodaghi, Ketra Schmitt, Member, IEEE, Pierre Watine, Benjamin C. M. Fung, Senior Member, IEEE

***Abstract*—Social media use has transformed communication and made social interaction more accessible. Public microblogs allow people to share and access news through existing and social media created social connections as well as access to public news sources. These benefits also create opportunities for the spread of false information. False information online can mislead people, decrease the benefits derived from social media and reduce trust in genuine news. We divide false information into two categories: unintentional false information, also known as misinformation; and intentionally false information, also known as disinformation and fake news. Given the increasing prevalence of misinformation, it is imperative to address its dissemination on social media platforms. This survey focuses on six key aspects related to misinformation: (1) clarify the definition of misinformation to differentiate it from intentional forms of false information; (2) categorize proposed approaches to manage misinformation into three types: detection, verification, and mitigation; (3) review the platforms and languages for which these techniques have been proposed and tested (4); describe the specific features that are considered in each category; (5) compare public datasets created to address misinformation and categorize into prelabeled content-only datasets and those including users and their connections; (6) survey fact-checking websites that can be used to verify the accuracy of information. This survey offers a comprehensive and unprecedented review of misinformation, integrating various methodological approaches, datasets, and content-based, user-based, and network-based approaches, which will undoubtedly benefit future research in this field.**

***Index Terms*— misinformation, rumor, satire, conspiracy theory, misinformation detection, misinformation verification, misinformation** mitigation

1. INTRODUCTION

SOCIAL media platforms have become the go-to source for news updates for a majority of people [1]. Many people have merged their everyday lives into popular online social sites, such as Facebook, Twitter, Sina Weibo, and Reddit, and rely on these microblogs as one of their primary news sources [2]. In other words, the advantages of social media, including ubiquity, accessibility, speed and ease for use, have made them indispensable sources of first-hand information [3]. The same factors that make social media an easily accessible source of information also make it an efficient vector for the creation and broadcast of false information about events in real-time [4].

Online false information has considerable offline consequences and poses a threat not only to platforms' users, but also to businesses, public health and governments. Examples include influencing election results [5], and false information regarding disease prevention. In one case, false information about the protective ability (and safety) of highly concentrated methanol to kill coronavirus infection resulted in the death of almost 800 people, and hospitalization of 5,876 [6]. As these examples illustrate, strategies are required to identify, counter and mitigate the propagation of online false information and reduce the negative impact.

Misinformation, as a primary form of false information, is typically shared or created without any malicious intent towards others [7], [8], [9]. Such misinformation commonly stems from misunderstandings, flawed representations, or cognitive biases caused by deficiencies in comprehension or attention [10]. To address the problem of misinformation, two primary methodological approaches have emerged: detection and spread minimization. The detection approach involves identifying false stories as they arise, particularly during breaking news, and developing systems to automatically verify the credibility of information and social media content [10]–[12]. Conversely, the spread minimization approach recognizes that misinformation is an inherent part of social media and proposes techniques to reduce its negative impact and minimize its further spread [13], [14].

Despite the growing concern over the prevalence of misinformation, there are still significant challenges to overcome. The current body of research on false information has predominantly focused on detecting intentionally false information [15]–[23], often overlooking the importance of verification and mitigation strategies to manage unintentional false information. Therefore, it is critical to explore and implement a range of techniques to address not only detection but also verification and mitigation of misinformation. Despite

Arezo Bodaghi and Pierre Watine are in the Concordia Institute for Information Systems Engineering. Ketra Schmitt is affiliaited with the Institute and the Centre for Engineering in Society. All three are at Concordia University, Montreal, Canada (E-mail: arezo.bodaghi@concordia.ca; ketra.schmitt@concordia.ca; pierre.watine@concordia.ca).

Benjamin C. M. Fung is with the School of Information Studies, McGill University, Montreal, Canada. (E-mail: ben.fung@mcgill.ca).

the growing concern over the prevalence of misinformation, the vast majority of published articles on this topic focus on detection and verification techniques. Fewer studies concentrate on mitigation techniques, which is a significant gap in the field. Additionally, existing approaches to misinformation management face persistent challenges due to the diverse contexts and structural differences between social media platforms. Distinguishing misinformation from intentionally false information is challenging, as the line between them is often blurred. To effectively combat misinformation, it is crucial to have a comprehensive understanding of its characteristics and variants. This understanding is necessary to identify and evaluate applicable techniques, which can then be assessed for their suitability and efficacy. In cases where suitable techniques are not available, new ones can be proposed to fill the gap.

This literature review paper aims to provide a comprehensive overview of techniques for detecting, verifying, and mitigating misinformation. By identifying limitations and gaps in current approaches, we aim to contribute to the development of more effective strategies. Our survey offers valuable insights and identifies areas for future research. Specifically, our contributions are:

- We provide a comprehensive review of detection, verification, and mitigation techniques aimed at addressing misinformation in its various forms including rumor, satire, and conspiracy theory.
- We review various approaches to address misinformation, beginning with early identification and tracking to ensure its accuracy, followed by verification or rejection. Additionally, we review proposed strategies for preventing the further dissemination of misinformation.
- We analyze the features and attributes of different methodologies proposed to tackle various forms of misinformation, including the models, platforms, and languages considered in the reviewed literature.
- We compile and categorize 53 public datasets representing the first comprehensive collection of both content-based and structure-based datasets on these topics. These datasets will be valuable for future research in this area.

The rest of the article is organized as follows: we first provide definitions for different types of misinformation as well as an overview of available literature reviews about false information in section 2. Then, Section 3 details our methodology in selecting papers, and Section 4 provides more detail about approaches for combating online misinformation along with public datasets. Section 5 discusses public datasets while Section 6 provides discussion and future works. Finally, Section 7 draws conclusions from the review.

2. PRELIMINARIES AND BACKGROUND

False information has evolved in meaning and usage over time. Generally speaking, false information refers to a news article or message published and spread through media, containing incorrect/fake information regardless of the motive and means by which it was transmitted [10]. False information spreads faster and deeper than true information, and tends to be sticky, persisting in memory [4], [24], [25].

There is no commonly accepted typology framework, no specific categorization criteria, nor explicit definitions to facilitate investigation of this issue. However, clear and common definitions of false information are crucial since the types of false information may require different theoretical analyses. A small number of literature reviews have attempted to characterize misinformation. According to [26], fake news can be classified into three types: serious manufacturing, large-scale farces, and humorous texts such as parodies and satires. In [27] authors distinguish fake news from deceptive news, misinformation, disinformation, false news, satire news, clickbait and rumors according to three criteria: authenticity, intention, and being news. Celliers and Hattingh [28] explored the motives behind spreading false information, leading to a description of false information types. Another important distinction to be characterized is between the terms, because some works use the terms interchangeably such as fake news and misinformation [29]–[32] or mistakenly like rumor and fake news [11]. Overall, the literature shows that false information can generally be classified as fake news, misinformation, and disinformation based on its facticity and intention [10], [33]–[36]. Another way to classify false information is to break it down into three elements [10], [37]: 1) The types of content being created and shared; 2) The motives that drive individuals to create and distribute this content; 3) The methods employed to disseminate this content.

According to a systematic review, disinformation comprises all forms of false, inaccurate, or misleading information intended to intentionally cause harm to the public or to generate profits [7], [18], [38], [39]. Fake news also refers to intentionally crafted, sensational, emotionally charged, misleading or totally fabricated information that mimics the form of mainstream news [33], [38], [40]. Fig. 1 illustrates different types of false information [9], [33], [34], [38], [41].

2.1 Definition of misinformation

Misinformation is defined as unintentional dissemination of false information that is misrepresented or misunderstood because of cognitive bias or omissions of pertinent data. Given the recent evolution of the term and the widespread misuse in various environments, we must first carefully define what should be considered as “misinformation". Some types of false information, such as accidentally misreported information, unverified rumors, implausible satire, and conspiracy theories fall into the category of misinformation [10], [42]–[44]. Considering rumor, satire, conspiracy, we define them respectively in the following subsections.

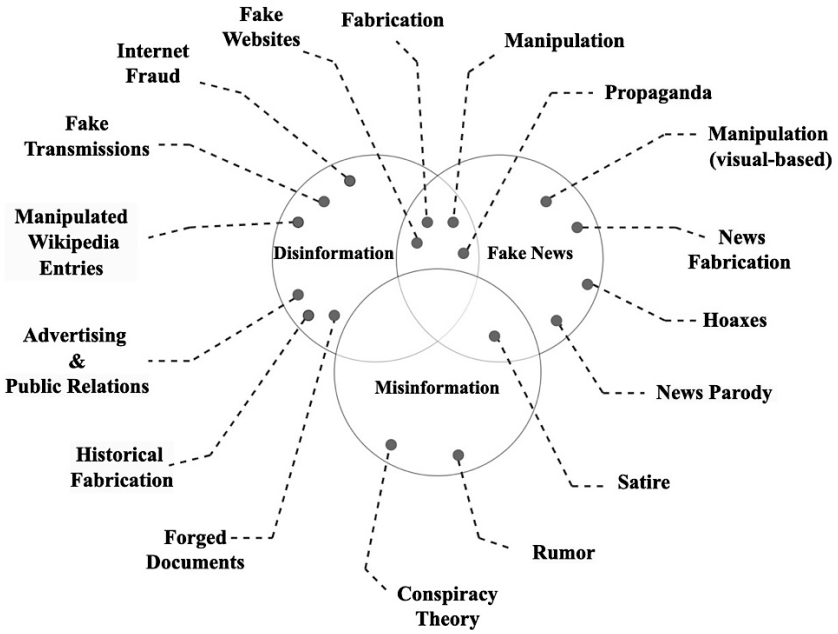

Fig. 1. The definition of different types of false information

2.2 Rumor

A rumor is commonly defined as an unverified statement of information and is often characterized by rapid spread [45], [46]. Rumors can be categorized into two types based on their lifespan: 1) New rumors that emerge during breaking news, 2) Long-standing rumors that persist and are shared over an extended period of time [11].

2.2.1 Rumors arise from Breaking News

These types of rumors are generally original and unique, i.e., have not been previously observed. A suspected and unconfirmed terrorist attack would be an example of a rumor arising from breaking news. This is a difficult category of rumor to detect as the names of actors, groups and locations vary with each instance. Hence new detection systems need to be designed with new vocabulary cases [47], [11]. Alternately, context clues could be detected based on previous breaking news rumors.

2.2.2 Long-Standing Rumors

Long-standing rumors persist for an extended period despite being disproved. Sometimes long-standing rumors persist despite overwhelming evidence to the contrary, and can be particularly resilient when they tap into pre-existing beliefs or emotions [47], [11].

2.3 Satire

In satire, news updates are presented through the use of humor or exaggeration [33]. These stories are presented as news that might be incorrect in fact, but the intention is not to deceive but rather to expose or identify shameful, corrupt, or otherwise ineffective behaviors. When satire is intended to mislead people, it falls into the category of fake news. Here, misinformation occurs with no intention to harm, but satire still has the power to mislead some people [10]. Compared to fake news, satire presents stories as news that are incorrect factually, but the intent is to call out, criticize, or expose behavior that is shameful, corrupt, or otherwise "bad " [48]. Legitimate news stories can have occasional factual errors, but these do not qualify as fake news.

2.4 Conspiracy

Conspiracy refers to a covert and managed scheme to bring about or prevent specific events [49] and often seeks to explain past occurrences as the result of the actions of a few organized actors [42], [50], [51]. Enders et al. found that social media serve as widespread channels for propagating conspiracy theories and misinformation by exposing large numbers of individuals to fringe concepts and ultimately finding credulous consumers of information [52]. Enders et al. found that frequent social media users were more likely to agree with conspiracy theories and misinformation.

2.5 Background

The number of published papers and proposed techniques investigating misinformation is increasing. We searched Scopus using the terms "online misinformation", "misinformation Detection\Verification\ Mitigation" and found 2806 papers between 2010-2022 (Fig. 2). This number is an underestimate as some articles use terms other than misinformation, such as fake news, etc. As seen in Fig. 2, the number of publications related to misinformation detection

is significantly higher than that of verification and mitigation. However, the number of papers on mitigation techniques as well as verification techniques experienced a gradual increase between 2018 and 2022.

In order to develop effective countermeasures, to misinformation, policy makers and developers must first estimate its magnitude [53]–[55]. The generation, impact, propagation, and management of misinformation have been studied from multiple perspectives, including computer science, sociology, journalism, and psychology. [56], [57]; which have led to the development of various tools, systems, and datasets to support research efforts [56], [58]–[62]. The main questions that have been identified are: Who are the main propagators of misinformation in the misinformation diffusion network, and what are its structural and dynamic characteristics? How can misinformation be reduced? In what circumstances and to what degree can misinformation be identified? What are the differences between the writing style, language of misinformation and correct information?

Zhou and Zafarani [27] reviewed techniques for detecting fake news from four perspectives: (1) Knowledge-based methods that verify if the knowledge in the news content (text) coincides with what is actually true; (2) Style-based methods, which analyze how fake news is written (e.g., if it is written using strong emotions); (3) propagation-based techniques that identify fake news by determining how it spreads online; and (4) source-based techniques that detect fake news by examining the credibility of news sources at different stages (when they are created, published online, and distributed via social media). They also considered all forms of false information including misinformation and disinformation, rumor to be fake news, but have distinguished between them according to three properties: authenticity, intention, and if it is news. Cao et al. adopt a different approach [63] by applying rumor detection from three perspectives: (1) handcrafted features based approaches, (2) Propagation-based Approaches, (3) Neural Networks Approaches. Collins et al. [64] categorized fake news into five different categories: Clickbait, Propaganda, Satire and Parody, Hoaxes, and others (Name-theft, framing, journalism deception) and classified detection techniques into eight classes: Expert or professional fact-checker, Crowdsourcing, Machine learning, Deep learning, Recommendation system, Hybrid technique, Expert-crowdsource, Graph-based method, Human-Machine approach. By comparison, Fernandez and Alani [57] defined four dimensions for combating online misinformation (rumors, false news, hoaxes, and elaborate conspiracy theory). They considered existing technological developments in four main research trajectories such as detecting, dynamics, validation and managing.

Sharma et al. [10] distinguish between unintentionally or intentionally false information and classify misinformation solutions between content-based and Feedback-Based detection techniques, which they defined as Intervention-Based Solutions. The application of deep learning techniques for detecting fake news, rumors, spam, false information, and disinformation was reviewed by [65]. The study highlights deep learning as a highly effective technique for social network data analysis and improving detection of misinformation, particularly in unlabeled and imbalanced data. Moreover, they identified a diverse range of challenges, including data quality, feature enrichment, federated inference, temporal modeling, data volume, and infrastructure limitations that need to be addressed for effective implementation.

Table 1 categorizes the techniques, approaches and targets of previous reviews. As this table shows, the majority of existing work focused on detection techniques, while mitigation techniques are rarely studied. As this overview demonstrates, previous reviews have considered misinformation and control techniques. However, this survey is the first to provide a comprehensive overview of all control techniques for handling misinformation.

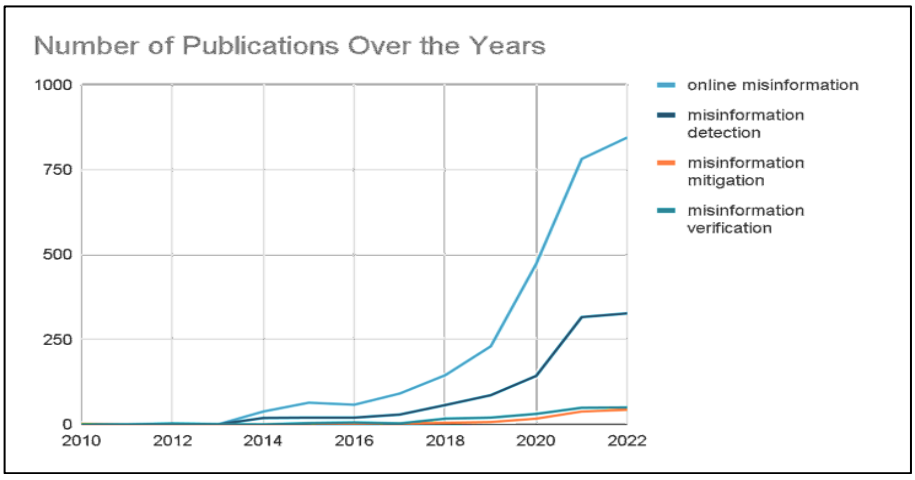

Fig. 2. Online misinformation publications

TABLE I
VARIOUS LITERATURE REVIEWS DISCUSS DIFFERENT TECHNIQUES FOR ADDRESSING ONLINE MISINFORMATION

Paper	Detection	Verification
[66]	1. Traditional models using handcrafted features 2. Deep Learning 3. Hybrid Machine Learning algorithm	1. Fact-checking websites 2. Traditional Machine Learning model 3. Deep Learning model
[67]	1. Content-based methods 2. Social Context-based Methods 3. Feature Fusion-based Methods 1. 4. Deep Learning-based Methods	1. Crowd Intelligence in Misinformation 1.1. Individual level 1.2. Crowd level
[27]	1. Knowledge-based (content) 2. Style-based (content) 3. Propagation-based 4. Source-based	1. Manual Fact-checking 1.1. Expert-based 1.2. Crowd-sourced 2. Automatic Fact-checking 2.1. Fact Extraction 2.2. Fact-checking
[68]	1. Source Detection 1.1. Single source detection 1.2. Multiple source detection	None
[63]	1. Handcrafted features 2. Propagation-based approach 3. Neural network-based approach	None
[69]	1. Classification approach 1.1. Machine Learning 1.2. Deep Learning 2. Other approaches 2.1. Retweet behaviour 2.2. Diffusion pattern 2.3. Anomaly detection 2.4. Hawks process	None
[70]	1. Source detection 2. Propagation models 2.1. Soft computing Models 2.2. Epidemiological Models 2.3. Mathematical Models	Fact-checking platforms

3. REVIEW METHODOLOGY

Based on the review above, we have identified the main research questions as follows:

Question 1*: What is the reported research in this field? What is the maturity level of the research*?

Question 2*: Which public datasets have been used?*

Question 3*: Which methods have been used for identifying misinformation?*

Question 4*: Which methods have been used for verifying misinformation?*

Question 5*: Which methods have been used for mitigating the spread of misinformation?*

Question 6*: Which types of misinformation has been investigated?*

Question 7*: What kind of features have been used?*

Question 8*: Which languages and platforms have been studied in this field?*

Question 9*: What are the advantages and disadvantages of proposed techniques?*

Question 10*: How effective are models trained on existing data for combating online misinformation over a long period of time?*

Based on these questions, we defined protocols for the review. We included the following electronic databases: Google

Scholar, Web of Science database, IEEE Xplore, Science Direct, arXiv, and ACM Digital Library, ACL Anthology, and Springer. These sources are chosen because of their comprehensive literature in this field. ArXiv is also included because some papers are published only at this open-access repository of preprints. To ensure a comprehensive review of the literature, we focused on publications from 2013 to 2023, as this is a rapidly evolving field with a large number of publications. In order to narrow down our selection, we excluded papers that were deemed too similar to other works or did not make a significant contribution to the field. We also defined the following inclusion criteria: 1. The main objective of the paper must be investigating ways to combat different types of misinformation.
2. Included items must be scholarly research.
3. Papers must be published in English.
4. The research must be published as a journal paper, conference paper, a book chapter or an arXiv paper.

We have conducted the search on the electronic sources listed above using the following strings:

- misinformation detection
- misinformation tracking
- misinformation veracity/stance classification
- mitigating the spread of misinformation
- combating/addressing misinformation

It should be noted that the term "misinformation" is often used to broadly refer to false information, including intentionally false information like fake news. To ensure that our selection criteria focused on unintentionally false information, we specifically chose papers that addressed this issue.

4. Managing Online Misinformation

We divide the methods applied into three types: detection, verification, and mitigation. Each category is further divided into subcategories based on the available methods (see Table II). As shown in Table I and supported by our own analysis, the majority of research in this area tends to focus on identification techniques, with some attention given to verification methods, while mitigation strategies are often overlooked.

4.1 Detection

In this section, we will review papers on detecting misinformation to address several questions. These include: which methods have been employed for identifying misinformation? What public datasets have been utilized? and what types of misinformation have been examined? Additionally, we will explore the languages and platforms that have been studied in this area, as well as the features that have been utilized.

Detection of misinformation aims to identify misleading claims using either an algorithm or trained artificial intelligence tools to classify the information [71]. Machine learning methods can serve as a classification tool for detection techniques. To train the artificial intelligence, some preferred to have a supervised approach with a pre-labeled training dataset [72] to differentiate between the misinformation from the various types of media (news articles and social media posts) while some use unsupervised deep learning to classify the misinformation into multiple categories to discover new features proper to the different types of misinformation [73]. Another branch of the literature focuses on the identification of the sources of misinformation [74] to simplify the misinformation detection protocols. One protocol proposes a two-step heuristic approach to find the most probable source [75]. This method can be extended to evaluate infected nodes and varied levels of confidence. Watine et al. [76] employ the Hawkes process to determine which social media community is influenced by which given a dataset of their posts. On a broader scale, this method could be used to evaluate how the information spreads on social media platforms and track down the source of misinformation and evaluate where and how it spreads. Once these sources are identified, it would be possible to tie them to blockchains that act as a certification stamp that can be seen by the browsers [77].

TABLE II

MANAGING MISINFORMATION: EFFECTIVE TECHNIQUES

Technique	Category
Detection	Content-based Features: Use the content itself, such as text, images, and videos and do not consider the users roles.
	User & Network Features: It detects misinformation by analyzing the role of users in spreading rumors as well as propagation patterns; how misinformation spreads among users. (user-based; network-based). Due to the fact that users alone cannot be used to detect misinformation, we consider all features relating to users and propagation structures in this category. All of them are in some way related to users.
	Hybrid techniques (Context and Content): It considers the content as well as social features and diffusion structure
Verification	Tracking: Relevant posts are gathered, and unrelated posts are filtered for potential misinformation
	Stance: assessing whether certain post support or contradict a claim
	Veracity: Predicting the veracity of misinformation
	Multi-task classification: Detecting, Tracking, Stance and Veracity classification. It detects misinformation first, then tracks them to verify their validity
Mitigation	User-based strategies: Identifying a set of users to broadcast counter messages
	Algorithm-based: Use an algorithm to pre-filter the information and send out the most suspicious pieces to experts for verification before they become viral
	Blockchain-based: Verify the primary sources of information using blockchains, facilitating the linkage between news and its sources

Through our quick or primary review, we were able to identify that there are six key factors that are essential for detecting misinformation. These factors include features,

detection models, platforms, languages, topics, and types of misinformation. However, to gain a more comprehensive understanding of these factors, we needed to conduct an in-depth analysis of relevant papers. Therefore, in this study, we attempt to extract these six factors from selected papers by carefully examining and analyzing the content. By doing so, we gain a deeper insight into the nature of misinformation and how it can be detected using these crucial factors.

4.1.1 Content-based strategies

User-Generated Content (UGC) is a term used to describe any content created by users, including texts, videos, images, reviews, live streams, and other forms of media. To identify misinformation within UGC, some techniques solely rely on the content-based features, including lexical, syntactic, and topical features, as well as writing styles [67], [78]. Classifiers can determine whether UGC is misinformation by using content-based features. This article explores current applications to identify the content-based features utilized in detecting various types of misinformation, including rumor, conspiracy, and satire. Table III summarizes some content-based features used in papers that proposed techniques for detecting misinformation solely based on content. By examining the different techniques and approaches employed by each paper, we can gain insight into the various features and methods used to identify different types of misinformation. The table also indicates the dataset, platform, language, and models used by each paper to detect specific types of misinformation.

According to Ye et al. [79] some conventional methods are mainly focused on feature engineering in dynamic and complex social media scenarios but fail to cover potential features in new scenarios as well as struggle to create elaborate interactions among significant features at a high level. In addition, Recurrent Neural Network (RNN) [80] based methods have also been shown to be unqualified for practical early detection of misinformation due to its bias towards the latest inputs. This led them to develop a novel method for detecting and classifying both truth and misinformation online early on. As a result of their approach, which is derived from Convolutional Neural Networks (CNNs) [81], A Convolutional Approach for Misinformation Identification (CAMI) can flexibly extract key features from a sequence of inputs and shape high-level interactions between those features, enabling effective identification of misinformation and practical early detection at the event level. It was unfortunate that they did not specifically point out any types of misinformation. They have validated the CAMI model using two large datasets from Twitter and Sina Weibo. In the case of a set of events, each event consists of a collection of microblog posts, and each microblog post has a timestamp. To automatically obtain key features of both misinformation and truth, an unsupervised method is used to learn the representation of input microblog posts, while a supervised method, CNN, is used to automatically obtain representations of input microblog posts. As compared to both conventional feature engineering methods and RNN methods, the novel approach was more effective. The model, however, is highly dependent on the training datasets as it performed differently on Twitter and Sina Weibo with accuracy rates of 93% and 73%, respectively.

In one study, a classifier was trained using a Generative Adversarial method without the need for a verified news dataset [82]. Combining Generative Adversarial Networks (GANs) [83] with Reinforcement Learning (RL) [84] algorithms resulted in high-quality and balanced representations of text for training. A key advantage of this technique is its explainable detection of rumors without the need for a verified news database; it also provides a powerful framework for identifying texture mutations; in addition, the model uses layered structures to avoid function mixture and to maximize performance. However, it falls short on explanations when it comes to identifying short sentences.

Conspiracy-related publications show that the least amount of attention is paid to this type of misinformation. In 2020, Serrano et al. [85] leveraged user comments to identify COVID-19 misinformation videos on YouTube using transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) [86], RoBERTa [87] that is a large-scale BERT. By extracting features from the first 100 comments on each YouTube video, they are able to detect misinformation more efficiently and more quickly than they would if they trained models on YouTube videos themselves. Labeling comments is easier and faster than labeling videos. They found that commentary on misinformation videos contains a significantly higher proportion of conspiratorial posts than on factual videos. This method has the limitations for example being highly dependent on comments, so that platforms should wait until each video gets an acceptable number of comments before training the classifier. Performing batch classification continuously in large-scale settings can be prohibitively expensive. For breaking news, this technique may not be very useful, since classifiers need to be updated.

The analysis of the results indicates that textual features are the primary contextual features utilized by content-based detection techniques, with limited works considering other content types such as images. Content-based features are mainly based on the number of words, characters, and sentiment features. Emotional features have been considered in some works, but stylistic and psychology features have received limited attention. Recent studies have employed deep learning techniques and pre-trained embeddings to extract content-based features. Most of the studies focused on rumor detection, while conspiracy and satire received less attention. Satire was often categorized as intentionally false news and excluded from analysis. Furthermore, Twitter was the dominant platform for analyzing user-generated content, and English was the most analyzed language. Overall, the analysis suggests that there is ample room for further investigation into the use of non-textual features, temporal modeling, and the detection of different types of misinformation beyond rumor.

TABLE III

SUMMARY OF CONTENT-BASED STRATEGIES FOR DETECTING MISINFORMATION

Paper	Content Features	Model	Platform /Language	Dataset/ Topic	Type
[88]	Textual feature: using Word2Vec # exclamation/question mark # words /characters; # positive/negative words # URL /@/Hashtags, Emoji, uppercase char; # shares; Sentiment score Visual feature: VGGNet (output of 2-19 layers)	Long Short-Term Memory (LSTM)	Twitter/ Weibo/ English	New Multimedia dataset (text and visual)	Rumor
[89]	Textual feature: Ratio capital letters Stylometric features (# words, Question mark, Punctuations, Hashtags , URL, @, Emoji) Emotional triggers features	Convolutional Neural Network (CNN)	Twitter/ English	PHEME	Rumor
[90]	It considers title and content while ignoring URL # Characters # Words # Sentences # DistilBERT tokens	DistilBERT	Twitter/ English	FakevsSatire; NELA-GT; Political news data Volkova_False news	Satire
[91]	37 Stylistic Features 10 Complexity Features 15 Psychology Features	One-way ANOVA test; Support Vector Machines (SVM)	English	Buzzfeed political news Burfoot & Baldwin	Satire
[92]	Emotional and semantic features extracted by model	Bidirectional Gated Recurrent Unit (Deep bi-GRU)	Sina Weibo/ Chinese	Weibo Dataset	Rumor
[93]	Only tweets with image are selected. Object Features of image extracted from ResNet that is trained on ImageNet Place and Scene Features extracted from ResNet that is trained on Places365 Hybrid Object and Scene Features extracted from ResNet that is trained on Places365 and ImageNet Image Sentiment extracted using CNN trained on VGG-19 Textual features extracted by fine tuning BERT	Vision and-Language BERT (ViLBERT) and SVM for classifying tweets	Twitter/ English	MediaEval/ Covid19	Conspiracy
[94]	Automatic feature selection	LSTM/ Introduce CallAtRumors	Twitter/ Sina Weibo/ English	Twitter and Weibo	Rumor

4.1.2 Social and Structure-based strategies

Structure-based strategies focus primarily on how misinformation spreads, as well as the importance of users in facilitating the spread of misinformation. Similarly, to the previous section, we review all factors that impact on detecting misinformation using social and Structure-based strategies. Some studies separated diffusion features from user-based features, but we consider them all to be related from our perspective. Any feature that considers users, their characteristics, and their role in propagating information as well as the structure of the network will be included in the category of Social & Structure features.

Most rumor detection methods rely on handcrafted features for implementing machine learning algorithms, which require tedious manual labor. However, some research has taken a different approach by detecting rumors through tracking changes in contextual information over time. Using RNN-based models Ma et al. [95] analyzed the change in contextual information of relevant posts over time in order to identify rumors at the event level. According to them, users dispute the truthfulness of a claim over time by posting various cues, resulting in long-distance dependencies. The results of experiments conducted on Twitter and Sina Weibo English data indicate that a RNN method is more efficient than rumor detection models that rely on handcrafted features. Moreover, Recurrent units and more hidden layers further improve the performance of the RNN-based algorithm, and it is faster and more accurate than existing techniques. On Twitter datasets, however, Support Vector Machines (SVM) [96] classifiers based on time-series structures continue to work better than RNNs.

In 2021, a participant-level framework for rumor detection was proposed (PLRD) [97]. Based on the diffusion network of the post, the PLRD model learns fine-grained user representations, including susceptibility, influence, user temporal, and integrates these features into a unique representation of rumor based on feature-level and user-level attention layers. With PLRD, the spread of rumor is considered as well as the influence of all those involved in the propagation process. Additionally, it detects those who initiate rumors as well as those who propagate them. Rumors can also be detected very early using PLRD. However, it does not take into account sentiment or content features, and only considers user-level features.

Early detection of misinformation, particularly rumor, is crucial for combating the spread and propagation of it. A multi-level early rumor detection model proposed by Nguyen et al. [98] evaluates potential rumors based on their abnormal behaviors, then they will be detected once they emerge. In the JUDO (Just-in-time rumor detection) model, the timing of the detection is crucial since timely detection can reduce the negative effects of rumor propagation, as already mentioned. In JUDO, anomaly scoring is done at two levels: first-order signals and high-order signals. When the former tracks anomaly signals individually, it incrementally computes an anomaly score for each element, whereas when the latter finds connected subgraphs with the maximum composite score, they will likely become rumors. Despite its advantages, this model has some limitations as well. It discards features associated with content for identifying rumors, for instance. It is also possible for the model to perform incorrectly when the user is only interested in rumors concerning a specific topic and is willing to sacrifice other topics to gain performance. By utilizing a topic-based filtering and monitoring tool, the model can be improved by exploring and aggregating topic information prior to lifting the data into the stream setting for rumor identification.

4.1.3 Hybrid techniques

In this section, we comprehensively review recent research papers proposing various innovative techniques for detecting online misinformation by considering content-based, user-based, and network-based features (see Table IV).

In 2013, to bridge the content semantics and propagation clues, a novel approach proposed for detecting rumors based on temporal, structural, and linguistic properties of rumor propagation [99]. To identify rumor or non-rumor topics based on tweets about the topic, they also built three classifiers using decision trees, random forests, and SVM. Later, a graph-kernel based hybrid SVM classifier was used by [100] that considered semantic features as well as propagation patterns. This hybrid method takes into account the internal graphical structure of messages, rather than focusing solely on lexical or semantic properties. As a result, it provides a flat summary of the propagation patterns of messages, such as the relation between reposts and their authors. Ma et al. proposed [101] a neural rumor detection approach based on recursive neural networks (RNNs). Based on recursive neural networks, they developed a bottom-up and top-down tree structure model to generate better integrated representations for rumor detection. Twitter datasets are used for experimental evaluation, which contain sets of widely spread source tweets as well as the propagation threads (replies and retweets). In spite of the fact that the proposed method is effective for detecting early stages of diffusion and it considers both structure and content semantics, it ignores the user's characteristics.

There have been several papers published in 2020 demonstrating the effectiveness of hybrid approaches in detecting rumors. A novel approach was proposed by [102] to detect rumor spread through the use of sequential classifiers and homophily assumptions. The first step is to build a friend network based on the follow-followers relationship. Afterwards, this network is encoded with the node2vec algorithm. To extract higher-level representations, the conversation structure is also used to process text information and user feature information. Finally, rumor detection is based on the fusion of information from all three components. With the use of this model, early detection of rumors is achievable with satisfying results. However, it is important to note that this approach was exclusively tested

on Twitter, and the classification is only applicable to tweets that have a high frequency of retweets.

In addition, Alkhodair et al. [47] introduced a method for detecting breaking news rumors about emerging topics. The study believes that emerging rumors can be perceived as true or false later, and do not necessarily have to be false when they are first detected. The proposed model considers both content and social features. A combined skip-gram-word2vec model is used to get word embedding from the training corpus. A Long Short-Term Memory (LSTM) [103] hidden layer is built by passing a sequence of vectors through a stack of weighted connections. When the LSTM model reaches its last time step, the predicted class is calculated as the output vector. With this model, breaking news rumors can be accurately identified based on a post's text alone. Over time, however, the model does not explicitly remember that emerging posts that are flagged as rumor could later be classified as non-rumors.

First work that focused on untrue rumors, which are used to denigrate, was done by Sangwan and Bhatia [104]. Denigration is a bullying tactic that destroys a person's reputation by spreading vulgar, cruel, derogatory, untrue rumors or mean. The most common use of this technique is to defame and discredit public figures including politicians and celebrities through rumorous stories, pictures, and videos. The purpose of this study was to propose a model for detecting potentially harmful posts (rumors) that denigrate bullying candidates. In order to confirm a case of denigration, the user profiles of these posts were examined. To uncover denigration, the model used three different categories of features such as text-based features, content-based features, and user-based features. Swarm-based wolf search algorithm was used to optimize the term frequency-inverse document frequency (TF-IDF) feature set, thereby reducing divergence and improving generalizability of the learning model. A classification model was built based on these optimal textual features as well as other content- and user-based features. There have been many studies using Twitter for experimentation, but this one used comment posted on Instagram photos as well as global celebrities' tweets and world leaders' tweets.

In 2021, using a feed forward neural network a propagation path aggregation model was proposed for rumor detection that integrates propagation structures and semantics of rumors rumors [105]. Rumor propagation is modeled as an independent set of paths, each representing a different context for talking about sources. By aggregating all paths, the propagation structure is represented. Furthermore, stance patterns in response propagation trees are captured using a neural topic model in Wasserstein autoencoder (WAE) framework without source posts. This model has the advantage of requiring fewer parameters and training quickly. Furthermore, the pre-trained neural topic model facilitates the use of unlabeled data in propagation path aggregation, especially when labeled samples are limited or rumors are spread early. User characteristics, however, are not considered in the proposed model.

An unified framework called ESODE was developed for rumor detection that integrates entity recognition, sentence reconfiguration, and ordinary differential equations [106]. Rumor texts are semantically analyzed using entity recognition. The next step is to reconfigure the sentence to improve the frequency of important words. Statistical features from three perspectives are collected in order to establish the complete feature map: linguistic features, characteristics of users involved in the propagation of rumors, and propagation network structures. Lastly, rumors are detected with the ordinary differential equation network (ODEnet). Besides considering linguistic features on the content of rumors, the proposed approach also takes into account characteristics of the users who propagate rumors, as well as the propagation network structures. In spite of the fact that this method includes user characteristics such as age and gender, it may not perform well if most users do not input a precise date of birth or gender.

Tu and his colleagues [107] developed Rumor2vec as a framework for rumor detection that combines text representations with propagation structures to identify rumors. A "union graph" that integrates multiple independent propagation trees was introduced first. Based on Twitter structure, the propagation tree can be modeled as a tweet cascade, which occurs when a tweet is published and then retweeted by another user. Two branches are included in the proposed framework: a text branch and a node branch. To extract high-order features from the transformed propagation sequence and the source tweet, CNN-based models are used in both branches. The final result represents a probability distribution over the corresponding set of classes. It is also noted that rumors posted by new users have no corresponding point in the union graph, and only their textual content is available. As a result, the model is reduced to one that uses only text.

Backward Compression Mapping Mechanism (BCMM) is another approach for early rumor detection [108]. It considers three categories of features as follows:
I) Textual features such as the representation of the entire post sentence and enhancing individual words and sentiments; II) Network-based features including the topology network attribution of the distribution graph of posters; III) Social features including the number of followers, following, replies and repost, etc. By combining BCMM with gated recurrent units (GRU), post content, topology networks, and metadata extracted from post datasets are represented. The model is highly effective and accurate at predicting early-stage events within a short period of time. Nevertheless, it only considers single-layer social network topology architectures, not multi-layer social network topology architectures (i.e., the relationship between multiple social network communities and online and offline networks).

TABLE IV

SUMMARY OF HYBRID-BASED STRATEGIES FOR DETECTING MISINFORMATION

Paper	Features			Model	Platform /Language	Dataset/ Topic	Type
	Content	User	Network				
[109]	Sentiment	# Followers # Favorites verified user	Poster Responder replies per user/ all replies	Graph Convolutional Network	Twitter/ English	Rumor Spreaders Dataset	Rumor
[110]	12 features (Fully presented in Table VI) Average text length Average sentiment score % of enquiry tweets	13 features (Fully presented in Table VI) Average # friends Average # followers Average # posts Average # reposts	6 features Fully presented in Table VI) max depth of propagation tree max depth of propagation tree / # tweets	Separable Conv LSTM SENet Gradient Boosting Decision Tree (GBDT)	Twitter/ English	PHEME	Rumor
[111]	Text Feature	User engagement Timestamp	Propagation features	linear and non-linear propagation (RDLNP)	Twitter/ English	PHEME Rumor-Eval	Rumor
[112]	word embedding to get word and phrase semantics. word co- occurrence relationship		Users and their interactions	k-Means BERT Louvain greedy community detection algorithm	Reddit/ 4Chan/ English	Covid19 conspiracy-theory	Conspiracy
[113]	Word Vectors: Wor2Vec Part of speech tags: # Occurrences of a certain POS tag in a tweet Ratio of capital letters Word Count -Question Mark Exclamation Mark Use of the period	# Tweets # Listed Count Follow Ratio Age Verified		Conditional Random Fields (CRF)	Twitter/ English	PHEME	Rumor
[114]	GloVe embeddings Subjectivity cues Psycholinguistic cues Moral foundation cues		user interactions if someone mentions (@) another user	LSTM, CNN	Twitter/ English	Volkova_ False news	Satire
[115]	GLoVe embeddings	No user identity	Post, Replies & Comments	Post level attention model	Twitter/ English	PHEME Twitter 15-16	Rumor

TABLE V

MISINFORMATION FACT-CHECKING/DEBUNKING SERVICES

Fact-checking service	Description\Topics Covered	Link
Factcheck	Promotes voter awareness and reduces deception and confusion in U.S. politics	https://www.factcheck.org/
Snopes	English Fact-checking Platforms; verifying and debunking urban legends	https://www.snopes.com/
Politifact	English Fact-checking Platforms; uses the "Truth-o-Meter" to rank the amount of truth in public persons' statements.	https://www.politifact.com/
Fullfact	A UK-based fact-checking organization that covers articles on the economy, health, education, crime, immigration, and law	https://fullfact.org/
Factcheckhub	An online English fact-checking website that combats misinformation, disinformation, hoaxes and rumors regarding a wide range of topics, such as the covid-19 pandemic, elections, the economy, health, security and governance.	https://factcheckhub.com/
Hoaxy	It visualizes and verifies the spread of claims on Twitter	https://hoaxy.osome.iu.edu/
Washington Post Fact Checker	Politicians rate statements	https://www.washingtonpost.com/news/fact-checker/
Gossip Cop	Is a fact-checked service in New York City. it rates 0-10 to each article	https://www.suggest.com/
Leadstories	The site contains English email rumors on politics, religion, nature, aviation, food, medicine, and many other topics	https://leadstories.com/
Mediabiasfactcheck	The mission of this independent website is to promote awareness of media bias and misinformation. Through a combination of objective and subjective measures, human evaluators determine the level of factual reporting and the bias of media sources.	https://mediabiasfactcheck.com/
Truth or fiction	The site contains English email rumors on politics, religion, nature, aviation, food, medicine, and many other topics. Originally focused on internet hoaxes and rumors, it has now expanded to include general fake news.	https://www.truthorfiction.com/
Scmp	global conversation about China	https://www.scmp.com/
Apnews	It covers U.S. News; World News; Politics; Sports; Entertainment; Business. Technology; Health; Science; Oddities; Lifestyle; Photography; Videos	https://apnews.com/
Tweet Cred	It is a real-time, web-based system for assessing the credibility of Twitter content	Chorome extension tool
Twitter Trails	Tracks the trustworthiness of Twitter stories	http://twittertrails.com/
Fatabyyano	Arabic fact-checking website	https://fatabyyano.net/
Misbar	Arabic fact-checking website	https://misbar.com/
AAP FactCheck	Australian Associated Press	https://www.aap.com.au/factcheck/
Rumorscanner	Bangladesh Associated Press	https://rumorscanner.com/
Décrypteurs	It is a Canadian website for fact-checking false information. It is in French/English	https://ici.radio-canada.ca/decrypteurs
Pagella Politica	Italian fact-checking service	https://pagellapolitica.it/
FactCheckNI	Northern Ireland's fact-checking service	https://factcheckni.org/
Truth Googles	The service provides fact-checking at the point of media consumption. Viewing content through various "lenses" of truth will give readers a more critical approach to even their most trusted sources.	https://www.media.mit.edu/projects/truth-goggles/overview/
FactWatcher	A variety of facts, such as situational facts, one-of-the-few facts, and prominent streaks, are considered	https://idir.uta.edu/factwatcher/

4.2 Verification

This section will concentrate on comprehensively reviewing academic papers related to the verification of misinformation. We will address several questions regarding this topic, such as: What methods have been employed to verify misinformation? What types of features have been utilized in this field? Which categories of misinformation have researchers investigated? What public datasets have been used in the study of misinformation? Which languages and platforms have researchers examined in this domain? What are the strengths and weaknesses of the proposed techniques for verifying misinformation?

An important step in dispelling misinformation is to verify and fact-check once it is detected. The time-consuming process of fact checking makes it nearly impossible to match the speed of social media [57]. Veracity assessment is intended to determine if a particular misinformation or rumor can be dismissed as false or true or whether it still requires investigation. Compared to other types of misinformation, it is mostly used for rumors. The Veracity Assessment has begun by [116] using real-world events. The author does not directly address the issue as a veracity assessment problem, but rather as a credibility assessment problem. In order to evaluate a message's credibility, in addition to message-based factors, topic-based factors, user-based features, and propagation-based factors were taken into account by the authors. In terms of message-based features, the authors considered two categories: Twitter-independent features (such as length, exclamation points, and sentiment words) and Twitter-dependent features (such as hashtag, re-tweet).

The verification process can be carried out manually or automatically. Fact-checking experts can verify claims manually, and there are fact-checking websites and systems that can be used [27]. A list of debunking/fact-checking tools is provided in Table V along with details. In addition, several fact-checked corpora were published, such as CREDBANK [61], Check-worthy [117], and RumorLens [118]. In order to create CREDBANK, more than 1 billion tweets were tracked in real time over a period of more than three months. CREDBANK is a collection of tweets, events, topics, and related human credibility judgements.

Manual fact-checking is not feasible due to the proliferation of information on social media. Scalability was addressed by developing automatic fact-checking techniques [119]. In a misinformation verification system, four tasks are generally performed: detecting, tracking, stance, and verifying. Typically, these tasks start with detecting unverified information and end with determining the estimated veracity value of the information; however, depending on requirements, some of these tasks may not be needed. For example, some works focus on a tracking [120], stance classification [121], verification [122], some others consider techniques that combine multiple tasks [73]. According to Zibizaga et al. [123] all four tasks were reviewed deeply, while in this section, we summarize papers describing at least one of the tasks or all of them for verifying online misinformation.

4.2.1 Misinformation Tracking

Tracking tasks collect and filter posts that discuss misinformation once it has been identified. By monitoring social media for posts discussing the misinformation, the tracking task is able to find postings that discuss the misinformation while eliminating irrelevant posts [123]. In other words, the output involves collected posts responding to misinformation rather than classifying it. The collected posts need to be labeled as either related or unrelated to a specific misinformation topic through annotation. [69].

The idea of using supervised machine learning to assess the relevance of new posts to detected rumors was first proposed by [45]. Despite this, the scientific literature does not contain much research on rumor tracking. Three types of features were considered by Qazvinian et al. for identifying rumors correctly (Network-based features, content-based features, and microblog-specific memes). Additionally, they published an annotated dataset with 10K tweets categorized as Rumor, Non-rumor, Believe, Deny/doubtful/neutral. Additionally, Hamidian and Diab [124] built decision trees based on data in the work done by Qazvinian et al. A number of pragmatic attributes were also added to Qazvinian's features, such as entities, events, sentiments, and emoticons. Recently, it has been proposed to use an ensemble model based on reinforcement learning to track rumors (RL-ERT), which aggregates multiple components, uses a weight-tuning policy network and exploits social characteristics to improve the performance [120]. The model demonstrates superiority, robustness, and effectiveness compared to other models.

Rumor tracking has not paid much attention to emerging rumors. Jaidka et al. [125] introduced a system called SocialStories based on incremental clustering to detect fine-grained stories within broader emerging topics on twitter. New text-based and time-based features were extracted using an incremental clustering method, that compares incoming tweets with existing stories and identifies emerging stories. An important contribution of this work is the development of text-based similarity calculation metrics, including an inverse cluster frequency similarity metric, as well as time-specific metrics that enable old entities to decay with time and maintain the homogeneity and freshness of stories.

4.2.2 Stance Classification

Posts associated with a misinformation are classified by their stance, indicating whether they support it, deny it, question it, or just comment (unrelated or unknown) on it. The veracity of a misinformation can be determined by the stance users have towards the misinformation; research has shown that misinformation that is greeted with more skepticism, such as denials and query responses, is more likely to turn out to be false later, while confirmed truths are generally greeted with affirmation [126].

Initially [45] annotated tweets as supporting, denying, or querying rumors. Later on, [127] suggested that the annotation scheme be expanded to four labels by adding an additional label, commenting. In misinformation verification, stance classification is the most difficult step [73]. Two reasons account for this: Four-way classification problems are inherently more difficult than binary classification problems, and imbalanced data makes them more difficult. Also, stance classification is more difficult than tracking. Taking into account the whole sentence is necessary when classifying stance. It is possible to classify posts together in the tracking task by filtering out obvious keywords, but stances are related to the semantic meaning of the whole sentence and are therefore more difficult to classify. An approach for detecting stances towards pre-chosen targets on Twitter has been introduced in SemEval-2016 task 6 [128]. Two tasks were formulated: Task A determines whether tweets are favoring, opposing, or neutral towards five targets (Atheism, Climate, Feminism, Hillary, Abortion). In Task B, participants are required to detect stances towards an unlabelled target while no training data is provided for this target.

4.2.3 Multi-task classification techniques

It is extremely challenging to determine whether each post makes a disputed factual claim or not. In some cases, misinformation may be detected accurately, while in others, it might be mistakenly identified as such. It is possible that some others remain unconfirmed. This process of automatically resolving misinformation can be broken down into smaller components known as pipelines, which include detecting, tracking, stance, and finally determining its veracity. By aggregating the evolving, collective judgments of users, these works believe that a classification system can assist track a misinformation's veracity status as it is exposed to this collective decision-making process [60]. The purpose of this section is to provide an overview of classification systems that bring together some of the components needed to create a system of this type. For verifying misinformation, previous sections focused on solo-task systems, but this section discusses multi-task systems.

Users can be warned that information in postings that may turn out to be false through a rumor detection system that detects posts whose veracity status is uncertain early on [129]. Liu et al. introduced Reuters Tracer, A system that detects and verifies news events on Twitter algorithmically in a timely manner [130]. Experimental results show that Reuters Tracer is able to uncover most breaking news stories faster than traditional Reuters reporting tools and most global media outlets. The system, for instance, detected the Brussels airport attacks 11 minutes after the first bomb went off. They said they were 8 minutes ahead of Reuters and Globally or locally, no media can match their speed.

For some multi-task classification techniques, each step in the process of rumor verification is developed as a separate component and then feeds into the next. On the other hand, [131] proposed a multi-task learning approach (2-3 tasks) that allows the main and auxiliary tasks can be trained together, resulting in improved rumor verification performance. Four different scenarios were used to assess the effectiveness of a multitask learning approach for rumor resolution: (A) Veracity classification using single task learning; (B) Combining stance and veracity classification to improve veracity classification; (C) Combining detection and verification tasks; (D) Combining all tasks: detection, stance and veracity. As a result of comparing two tasks from the verification pipeline, they found that joint learning outperforms single-learning rumor verification. Performance is further improved when all three tasks are combined.

An integrated framework for rumor detection and stance classification has been proposed by [132]. They used deep neural networks to train both tasks jointly while each task retaining the ability to learn task-specific features independently. For stance, they considered four classes, including Support and Deny, Question, and Comment. Using news reports and real-world tweets, the experiment has demonstrated that compared with many strong baselines, the multitask approach consistently outperforms them in both tasks, suggesting a better strategy than training these rumor-related tasks individually with a multi-task architecture.

There was also a multi-task learning approach for detection and stance classification of rumors published in 2019 [133] In this model, user credibility information is incorporated into the rumor detection layer, and unlike the approach proposed by [101], the attention mechanisms are used for rumor detection process. They derived the credibility information from various user profile features, such as: is it a verified account? Does the profile include location information? Does the profile have a detailed description?

A Variational Autoencoder-aided Multi-task Rumor Classifier (VRoC) proposed by Cheng et al. [73] that combines all four components (detector, tracker, stance, veracity) for rumor verification. The VRoC system uses a variational autoencoder to create rumor classification at the tweet level. It uses Bidirectional-LSTM (Bi-LSTM) [134] for classifying and it is able to classify previously seen or unseen rumors. Furthermore, it is efficient in terms of parallel computing and rumor detection speed.

4.3 Mitigation

In spite of an extensive research effort devoted to detecting misinformation and creating automated systems for checking credibility and verifying social media content, online misinformation spreads like wildfire [14]. Numerous scholars hold the opinion that misinformation will always exist on social media, and therefore, their focus is on reducing its negative impact rather than eliminating it entirely. This section will examine various techniques and their attributes proposed for minimizing the further spread of misinformation on social media. There are various strategies that can be employed to limit the spread of misinformation, including disseminating debunking and counter messages

through users, analyzing the patterns of information dissemination, and exploring innovative techniques for verifying primary sources of information. It is crucial to adopt such measures to effectively address the persistent challenge of online misinformation.

4.3.1 Debunking strategies

For misinformation mitigation, one strategy is to categorize users based on their beliefs and scientific knowledge about the rumors and isolate posts from those who spread the rumors [135]. The main problem here is that the user's opinions may vary depending on their social neighbors. Another way to combat misinformation would be to find people who are willing to broadcast counter messages to limit the number of people believing it [14]. It would be best if these people were experts in the field to ensure the message is conveyed clearly and convincingly.

4.3.2 Algorithm-based strategies

An alternative approach suggested in a paper is to employ the algorithm called (curb) to detect problematic content before it is posted on social media. While this method proves effective as an initial barrier, it is not comprehensive on its own and may generate numerous false negatives. [136]. On this same idea of banning content, there is also a proposition by Marco Amoruso that involves a two-step heuristic approach to first identify the misinformation and then monitor its spreading nodes in the network. This approach is very scalable with the size of the network under study, but it was not tested on cases where the spreading nodes were only identified with a certain level of confidence [75].

4.3.3 Blockchain-based strategies

Blockchain technology is an emerging approach to combating misinformation by providing a certification mechanism for verifying the authenticity of primary sources of information [77]. Although the absence of a blockchain certification does not necessarily imply that the source is providing false information, the presence of a certification blockchain could help in identifying false information. This is because blockchain technology provides a secure and tamper-proof record of the original source, making it difficult to alter or manipulate the information.

However, the use of certification blockchains also raises several concerns. One concern is who would be responsible for generating these certification blockchains, and how can we ensure that the individuals or organizations entrusted with this task can be trusted? Another concern is the possibility of individuals forging valid blockchains to certify false information. It's important to note that the effectiveness of using certification blockchains in combating misinformation is still hypothetical and has not been measured by any metrics. Thus, it's crucial to either apply these suggestions or find a new method that is easy to apply in order to mitigate the spread of misinformation.

4.4 All considered features

In this section, we present a comprehensive list of features that have been identified in the literature as useful for addressing online misinformation. The features have been classified into distinct groups, including content, user, and network, and can be implemented through a variety of Detection, Verification, and Mitigation techniques (as outlined in Table VI). It should be noted that while previous studies have also explored features for identifying misinformation, our survey has covered a significantly larger number of features compared to those studies (e.g., [66], [137]). This emphasizes the criticality of employing a more comprehensive and nuanced approach to address the problem of online misinformation effectively.

4.5 Platforms

Social media platforms have a significant impact on the dissemination and management of misinformation. However, due to variations in user demographics, platform structures, content, objectives, anti-misinformation strategies, data accessibility, and popularity across different regions and among diverse audiences, proposed techniques, and interventions for addressing online misinformation may not be universally applicable. The purpose of this section is to identify the most frequently used social media platforms in academic research, in order to assess the level of attention that each platform has received. This analysis can help identify which platforms have been understudied and which have been more closely examined, enabling researchers to better understand potential gaps in the literature and identify areas of interest for future investigations.

Researchers can access social media platforms, collect data, and store it by utilizing application programming interfaces (APIs). [138]. The Twitter platform is frequently used due to its public nature by default and accessibility of users' profiles by non-users as well. A Twitter API[1] enables researchers to retrieve public Twitter data that users choose to share with the world. APIs are also available for Facebook[2], Sian Weibo[3] and Reddit[4] that make their data accessible. Facebook has almost the most users (2.9 Billion5), but collecting data is challenging due to its fenced-off nature. While some Facebook profiles are publicly accessible, many users use privacy settings to restrict access. According to the distribution of research findings shown in Fig. 3, Twitter is the most widely studied social media platform, followed by Sina Weibo and YouTube. In contrast, Reddit and Instagram have received comparatively little attention. Some studies have also considered WhatsApp [139] and Telegram [122], [140].

[1] https://developer.twitter.com/en/docs/twitter-api
[2] https://developers.facebook.com/docs/
[3] https://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3_V2/en
[4] https://www.reddit.com/dev/api/

TABLE VI

FEATURES CONSIDERED FOR COMBATING MISINFORMATION

Level	Features
Content	Time-orientation (Verb tenses)
	Topic
	Writing style
	Formality & Sophistication of posts
	Inferring & tentative words
	Opinion & insight words
	Punctuation
	Likes
	Comments
	Shares
	Sentiment score; Sentiment word
	# Mentions
	Text length
	% Enquiry posts
	% Posts with hashtags
	% Posts with question marks
	% Posts with exclamation marks
	% Posts with pictures or videos
	% Posts with URL
	Word and phrase semantics
	Verb quantity
	Word length
	Character Length
	% Pronouns; # different pronouns
	# Uppercase characters
	Contains of Emoji
	Positive/negative words
	Ratio of capital letters
	Use of Period
	# Emotion words
	# Stop words
	# Quotes
	# Interrogatives
	# Comparison words
	# Adverbs/ Adjectives
	# Causal words
	# Discrepancy word
	# Tentative words
	# Certainty words
	# Differentiation words
	# Affiliation words
	# Risk words
	Visual Features
	Hashtags
	Source tracking

Level	Features
User	User originality
	Verified users
	Description or bio
	Personal picture; Profile Logo
	Gender
	Age
	Education level
	User Controversiality
	Role
	# Follower
	# Following
	User influence (# follower / # following)
	# Posts
	# Reposts
	Favorites
	Engagement
	# Lists a user belongs to
	Time between the first post & registration
	Average time interval between posts
	Average days that a user's account exists
	# Posts per day
	# Comments
	Days since registration for influential accounts
	Location
	Virality
	Source credibility
Network	Network Diameter
	Low-to-high diffusion fraction
	% Nodes in largest connected component
	Ratio of depth to breadth on average
	New user ratio
	Original posts ratio
	% Posts including links to outside sites
	% Isolated nodes
	Max depth of propagation
	Max depth of propagation / # Posts
	# Leaf nodes / # Posts with responsive
	Source-tweet timestamp
	Timestamp of reactions to source tweets
	In-degree
	Out-degree
	Clustering coefficient
	Key nodes
	User's role in the network
	Interaction patterns

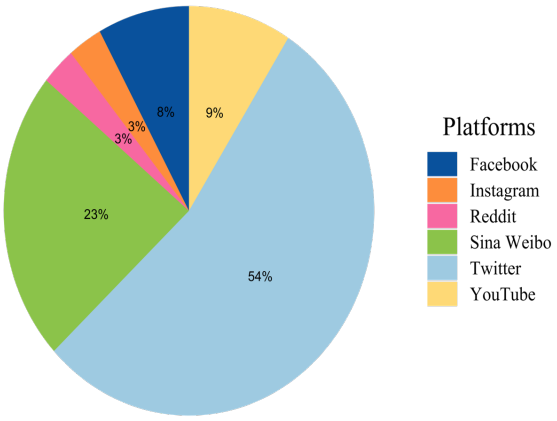

Fig. 3. Platform Distribution in Misinformation Research

5. PUBLIC DATASETS

Datasets are an essential component of addressing misinformation through a variety of techniques, particularly in machine learning and deep learning-based approaches that rely on high-quality training data. However, obtaining data from all social media platforms is a complex task, and the effectiveness of a dataset from one platform may not necessarily transfer to another due to differences in structure, content, users, and techniques. Additionally, some datasets may only contain post content, while others may incorporate social and network-based features, creating further variations in the types of data available for analysis.

Language is also an essential factor to consider when addressing online misinformation, particularly in content-based models. However, the majority of existing methods have only been tested on English-language content, limiting the scope of their applicability. To address this gap, our survey differs from related surveys in several ways. First, we discuss most of the publicly available datasets for detecting, verifying, and mitigating misinformation, rumor, conspiracy theories, and satires. Second, we have separated datasets that contain only content from those that include propagation and social features, providing a more detailed analysis of the available data. Third, our survey includes 53 datasets, making it the most comprehensive source of available datasets. Finally, for all datasets, we provide links to public repositories, language, multimodal type, size, a short description, as well as labels for those that are annotated.

A study by D'Lizia and colleagues in 2021 [141] reviewed and evaluated datasets for detecting fake news, which they defined as "all forms of inaccurate, misleading, or false information, designed, presented, or promoted with an intent to harm the public or to profit from it," including rumors, satire, and conspiracy theories. Our survey is one of the richest surveys on misinformation's datasets as we reviewed 53 datasets into two different categories (content and structure-based), making our study one of the most comprehensive to date (see Tables VII-VIII). Our survey fills a crucial gap in the literature, providing researchers and practitioners with a more extensive understanding of the available datasets for addressing misinformation and highlighting the importance of considering language and the types of data included in these datasets for effective analysis.

The PHEME dataset is a widely used dataset for research on rumor detection. The dataset was developed in a study by Zubiaga et al. [113], which aimed to create a multilingual dataset to facilitate research on rumor detection in different languages. The dataset includes 4,842 tweets from 330 rumor threads, covering nine different newsworthy events. An additional investigation on the PHEME dataset examined five breaking news events and included 1,972 rumors and 3,830 non-rumors [142]. The availability of the PHEME dataset has allowed researchers to develop and test different approaches for detecting rumors in social media. By including both true and false rumors, as well as unverified information, the dataset provides a realistic representation of the types of information that can be found on social media. Furthermore, the inclusion of multiple languages allows researchers to investigate the effectiveness of different methods across languages, which is particularly important given the global nature of social media.

RumorEval is another dataset used for detecting and verifying rumors. It was introduced as part of task 8 of the SemEval-2017 competition [143] and consists of 325 tweet threads. Later, it was upgraded to RumorEval 2019 [144] which includes English, Russian, and Danish tweets annotated for the detection and verification of rumors.

The PHEME and RumorEval datasets share a number of notable events, such as the Ferguson unrest, Sydney hostage situation, Ottawa shooting, and Germanwings plane crash. However, they differ in their annotation levels. PHEME provides three levels of annotation: firstly, threads are classified as either rumors or non-rumors; secondly, rumors are categorized as true, false, or unverified; and finally, stance classification at the tweet level is performed using crowd-sourced annotations on a subset of threads used in RumorEval. On the other hand, RumorEval offers annotations at both the thread and tweet levels and is available in English, Russian, and Danish. The PHEME dataset, available in English and German, provides annotations solely at the thread level. This variation in annotation types and languages offers researchers the opportunity to evaluate different rumor detection and mitigation algorithms across a wide range of languages.

Overall, the availability of these datasets has significantly improved research in the area of rumor detection, verification, and mitigation, and they continue to serve as important resources for researchers in this field.

6. Discussion and Future Work

After conducting a comprehensive review of the available literature, we have pinpointed several noteworthy shortcomings in the existing systems designed to tackle online misinformation. These limitations include:

Heavy reliance of detection techniques on training datasets: Detection systems based on machine learning and deep learning techniques are highly dependent on training datasets, which may not be effective in coping with breaking news or diverse types of misinformation. There is a need for specific and precise datasets that are annotated by experts in the related domain, which is challenging to obtain in real-time breaking news situations.

Diversity: Social media platforms are used by people from diverse backgrounds, languages, and cultures, and misinformation can be targeted at specific communities or regions. It requires a deep understanding of local contexts and specific datasets to effectively tackle misinformation, which may not be readily available in the public domain. Moreover, no multilingual annotated dataset which is applicable for most languages exists.

Various types of misinformation: Many public datasets do not have a precise definition or explicit difference between various types of false information. They may broadly categorize content into misinformation or fake news, which is not effective in detecting different types of misinformation.

Bias: Algorithms used for early detection of misinformation may inadvertently amplify it, highlighting the need for tracking and verification of misinformation.

Then, there is a need for tracking, verification of misinformation, but it also has own limitations. Effective verification of misinformation requires tracking its diffusion and monitoring user performance in disseminating news. However, respecting user privacy can make it challenging to track the source of misinformation and identify individuals responsible for spreading it. Moreover, even with detection and verification techniques, misinformation may still remain and erode trust in institutions, media, and experts. Mitigation techniques are required to counter them and address their impact on society.

Overall, intervention-based solutions have higher efficacy, but the models are more difficult to fit, requiring more data and are applicable in fewer situations. We conclude that a multitask technique is required for detecting, verifying and then mitigating propagation over social media platforms. We also identified persistent challenges to successfully identifying a set of best approaches for mitigation and prevention. Chief among these is the lack of a topic-sorted and labeled standard dataset that can help to identify off-topic misinformation. In addition, those techniques that only consider content-based or user-based or network-based features are less successful and reliable than those based on hybrid features techniques. In addition, many of the techniques suggested operate solely on text-based content, despite the fact that social media platforms also contain shared content in various other formats, such as images, audio, and video. The vast majority of published articles on misinformation are aimed at detection and verification, while a few are aimed at mitigation. A future survey can be conducted specifically to review the characteristics of mitigation-based techniques. Existing approaches to misinformation management are further stymied by structural differences between social media platforms. These differences make developing common solutions to mitigating false information using the proposed techniques challenging.

7. Conclusion

Through our six-step process, we were able to gain a comprehensive understanding of the current state of the field and identify potential avenues for future research: (1) we distinguished between intentionally false information and unintentional false information, which was critical for our study as it allowed us to categorize different types of misinformation accurately. This distinction helped us to identify the various forms of misinformation, including rumors, conspiracy theories, and satire; (2) we explored three perspectives on how to address online misinformation, including detecting, verifying, and mitigating. This allowed us to identify potential strategies and techniques for combating misinformation in different contexts. Our research showed that the most effective approach to combating misinformation involves a combination of these three perspectives; (3) we focused on identifying the most effective fact-checking services available for different languages and countries. Our research found that these services vary significantly in their effectiveness and that different services are more effective for different languages and countries; (4) we compiled and categorized 53 publicly available datasets that could be used to combat misinformation. These datasets can be used to train machine learning models that can automatically detect misinformation; (5) we analyzed the strengths and weaknesses of existing works on the topic. Our investigation uncovered the continued presence of considerable limitations. Despite the numerous methods proposed to combat misinformation, our findings indicate that they remain inadequate in fully addressing the issue; (6) finally, we identified the available features that could be integrated into approaches aimed at combating misinformation. Our research showed that these features include social network analysis, user modeling, and sentiment analysis.

Overall, our comprehensive review of the literature on misinformation provides valuable insights into the current state of the field and identifies areas for future research and development. We hope that our work will contribute to the development of more effective and efficient strategies for tackling online misinformation.

TABLE VII

A SUMMARY OF VARIOUS ANNOTATED/LABELED DATASETS FOR MISINFORMATION (CONTENT)

Dataset	Data Size	Type / Language	Label	Platform	Link	Source
LOCO	88-million word language of conspiracy corpus	Text / English	Conspiracy	Conspiracy websites	https://osf.io/snpcg/.	[145]
FEVEROUS	87,026 verified claims	Text / English	Supported; Refuted; Not enough info	Wikipedia	https://fever.ai/dataset/adversarial.html	[146]
Twitter Fake News	401,414 fake news, 290,627 real news	Text / English	Fake or Real	Twitter	https://github.com/s-helm/TwitterFakeNews	[147]
ArCOV19-Rumor	1M COVID-19 tweets. 138 verified claims from fact-checking websites; 9.4K relevant tweets identified & annotated.	Text / Arabic	False; True; Other	Twitter	https://gitlab.com/bigirqu/ArCOV-19/ -/tree/master/ArCOV19-Rumors	[148]
ArCorona	30M collected tweets related to COVID-19; 8K annotated tweets	Text / Arabic	13 classes like rumor, info, advice	Twitter	https://alt.qcri.org/resources/ArCorona. tsv.	[149]
Arabic COVID19 misinformation	4.5M tweets; 8.8K annotated;	Text / Arabic	Misinformation or No	Twitter	https://github.com/SarahAlqurashi/COVID-19-Arabic-Tweets-Dataset	[150]
Covid19-Twitter	490M tweets; most frequently used terms	ID[1]/ multilingual[2]	None	Twitter	https://github.com/thepanacealab/covid19_twitter	[151]
Emerging rumor	34 hashtags using #WhereAreTheChildren	Text / English	rumor; Non_rumor	Twitter	http://dmas.lab.mcgill.ca/data/RumorsNonRumorsCaseStudyData.zip	[47]
ClaimBuster	23,533 statements	Text / English	Non-factual ; Unimportant factual; Check-worthy factual	U.S. presidential election	https://zenodo.org/record/3609356#.Y4E4N-zMJAc	[152]
Coronavirus (COVID-19) Tweets Dataset	310M tweet IDs and sentiment scores	ID / English	Positive: (0,+ 1] Negative: [-1, 0)	Twitter	https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset	[153]
FEVER 2	1174 claims; Adversarial Attacks	Text / English	Supported; Refuted; Not enough info	Wikipedia	https://fever.ai/dataset/adversarial.html	[154]
NELA-GT-2018	713k articles collected from 194 news & media outlets like hyper-partisan, mainstream, conspiracy sources	Text / English	Veracity' s dimensions		https: //doi.org/10.7910/DVN/ULHLCB	[155]

[1] Using the Tweet IDs, you can query Twitter's API to retrieve the entire Tweet object, including tweet content and the author's metadata.

[2] English, French, Spanish, German, Russian

TABLE VII (CONTINUED)

Dataset	Data Size	Type / Language	Label	Platform	Link	Source
rumor Dataset	3 million users, 4 million tweets, 305115 linked articles, 28893 hashtags, around 1022 rumors	Text / English	True; False; Mixture; Mostly True; Mostly False; Unproven	Twitter Snope	http://tiny.cc/p1s2qy	[98], [156]
FEVER	185,445 claims	Text / English	Supported; Refuted; Not enough info	Wikipedia	https://fever.ai/dataset/fever.html	[157]
PHEME dataset for Rumor Detection and Veracity Classification	6,425 Threads; 4,023 non-rumors; 2,402 rumors, with 1,067 true rumors, 638 false rumors, and 697 unverified rumors;	Text / English	Rumor; Nun-rumor Each rumor is also annotated with its veracity value: True, False or Unverified	Twitter	https://figshare.com/articles/dataset/PHEME_dataset_for_rumor_Detection_and_Veracity_Classification/6392078	[131]
FakevsSatire	283 fake news articles & 203 satirical stories; The article titles, links, and full texts are included along with a thematic analysis, identifying significant themes such as hyperbole, conspiracy theories, racism, and discrediting reputable sources.	Text / English	Satire Fake news	Articles focused on American politics	https://github.com/jgolbeck/fakenews	[48]
SemEval (RumorEval) 2019	rumorEval 2017 data is the training data in 2019	Text / English / Russian / Danish	stance detection rumor detection; true/false	Reddit Twitter	https://figshare.com/articles/dataset/rumorEval_2019_data/8845580	[144]
Science	Including retweet cascades which are linked to rumors.	Text / English	True; False; Mixed	Twitter	https://goo.gl/forms/AKIlZujpexhN7fY33	[4]
PHEME-5 Experiment over five news datasets	5,802 tweets, ,972 rumors and 3,830 non-rumors	Text / English	rumor; Non-rumor	Twitter	https://figshare.com/articles/dataset/PHEME_dataset_of_rumors_and_non-rumors/4010619	[142]
SemEval (rumorEval) 2017	4,519 tweets; 297 source tweets; 7100 discussion tweets	Text / English	True; False	Twitter	https://alt.qcri.org/semeval2017/task8/index.php?id=data-and-tools	[143]
Rumor Detection over Varying Time Windows	rumors between 2006 and 2009; 60 rumors and 51 non-rumors	Text / English	Rumors; non-rumors	Twitter	https://figshare.com/articles/dataset/Rumor_Detection_over_Varying_Time_Windows/4550245	[158]
Political news dataset	75 stories per category	Text / English	Real, fake, satire	News Source	https://github.com/rpitrust/fakenewsdata1	[91]
Volkova_False news	Propaganda: 30,894 Satire: 2083 Hoax: 1341, clickbait: 836	Text / English	Propaganda, Hoax Clickbait, Satire, verified	Twitter	https://aclanthology.org/P17-2102/	[114]

TABLE VII (CONTINUED)

Dataset	Data Size	Type / Language	Label	Platform	Link	Source
Twitter 15	374 non_rumors; 370 false rumors 372 true_rumors; 374 unverified rumors 276,663 users	Text / English	Non_rumor; True_rumor; False_rumor; Unverified_rumor	Twitter	https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0	[159]
Twitter 16	205 non_rumors; 205 false rumors 205 true_rumors; 203 unverified rumors 173,487 users	Text / English	Non_rumor; True_rumor; unverified rumor	Twitter	https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0	[159]
PHEME	330 Threads 4842 tweets.	Text / English/ German	True_rumor; False_rumor; Unverified_rumor	Twitter	https://figshare.com/articles/dataset/PHEME_rumor_scheme_dataset_journalism_use_case/2068650	[160]
Emergent	300 rumored claims; 2,595 associated news articles For stance classification	Text / English	supporting; Against refuting; Observing reporting	Twitter Snopes	https://github.com/willferreira/mscproject	[161]
Twitter and Sina Weibo datasets	Twitter: 498 rumors and 494 non-rumors Sina Weibo: 2313 rumors and 2351non-rumors	Text / English	Rumor; Non-rumor	Twitter Sina Weibo	http://alt.qcri.org/~wgao/data/rumdect.zip	[95]
SemEval-2016	4870 English tweets regarding six commonly known targets in the United States	Text / English	Favor; Against; Neither	Twitter	https://alt.qcri.org/semeval2016/task6/index.php?id=data-and-tools	[128]
Sina Weibo dataset	2601 false rumors, 2536 normal messages and with 4 million distinct users involved in these messages	Text / Chinese	False rumors with at least 100 reposts; normal rumor	Sina Weibo	http://adapt.seiee.sjtu.edu.cn/~kzhu/rumor/	[100]
CREDBANK	60M tweets annotated by 30 human credibility	Text / English	Certainly Inaccurate; Probably Inaccurate Uncertain (Doubtful); Probably Accurate; Certainly Accurate	Twitter	https://github.com/compsocial/CREDBANK-data	[61]
Rumor has it	10K manually annotated tweets for rumor verification	Text / English	Rumor; Non-rumor; Believe; deny/doubtful/neutral	Twitter	https://github.com/vahedq/rumors	[45]
Satire Corpus	4000 newswire documents and 233 satire news articles	Text / English	True; Satire	newswire from English Gigaword Corpus	http://www.csse.unimelb.edu.au/	[162]

TABLE VIII

A SUMMARY OF VARIOUS DATASETS INCLUDING USER AND NETWORK-BASED FEATURES

Dataset	Data Type and Size	Platform / Label	Link	Source
Lastfm-social	Node: LastFM users from Asian countries (7624); Edges: relationships between users (27,806)	Last.FM	https://snap.stanford.edu/	[72]
Deezer-social	Node: Deezer users from European countries (28,281); Edges: relationships between users (92,752)	Deezer.com	https://snap.stanford.edu/	[72]
ArCOV-19	2.7M Arabic tweets by over 690k unique users; Retweets: 7,925,821; Replies: 1,476,950	Twitter / Fake News, Misinformation	https://paperswithcode.com/dataset/arcov-19	[148]
CHECKED	fact-checked Chinese dataset of social media for COVID-19 2,104 verified microblogs; 1,868,175 reposts; 1,185,702 comments; 56,852,736 likes	ground-truth labels, textual, visual, temporal, network data	https://github.com/cyang03/CHECKED	[163]
COVID-19 conspiracy theory	99,039 Tweet IDs with #FilmYourHospital; Node: 41,903 unique user; Edges: represent interactions among users (reply, retweet, or mention) 79,736	Twitter / Misinformation, Infodemic, Conspiracy Theory	https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP2/BSGQGS	[164]
Rumor collection	4 M tweets; 3 M users; 28893 hashtags; 305115 linked articles; almost 1022 rumors for 6 months	Twitter & Snopes as a rumor-debunking / Rumor	http://tiny.cc/p1s2qy	[98], [156]
Advogato	Nodes: Advogato users; directed edges: trust relationships (certification) Edge weights: certifications as apprentice (0.6), journeyer (0.8) and master (1.0)	Advogato	https://figshare.com/articles/dataset/advogato_tar_bz2/3082042/1	[165, p. 2]
Digg Friends	Nodes: Digg users; Edges: friendship links	Digg	https://figshare.com/articles/dataset/Digg_2009_social_news_votes_and_graph/2062467/1	[166]
Youtube links	Nodes: Youtube users Unweighted Edges: User Links	Youtube	https://networkrepository.com/youtube-links.php	[167]

TABLE VIII (CONTINUED)

Dataset	Data Type	Platform / Label	Link	Source
Political Blogs	Nodes: political blogs about politics in the United States of America directed Edges: hyperlinks between nodes		http://konect.cc/networks/dimacs10-polblogs/	[168], [169]
KAIST	102 topics, each contain at least 60 tweets; 47 rumors; 55 non-rumors Nodes: rumor spreaders; Edges: information diffusion	Twitter / Rumor	http://mia.kaist.ac.kr/publications/rumor	[99]
Facebook lists	Nodes: Facebook users; Edges: Friends lists	Facebook	http://snap.stanford.edu/data/egonets-Facebook.html	[170]
Twitter lists	Nodes: Twitter users; Edges: Friends lists	Twitter	http://snap.stanford.edu/data/egonets-Twitter.html	[170]
YouTube friendship	Nodes: YouTube users; Edges: Members of groups	YouTube	https://snap.stanford.edu/data/com-Youtube.html	[171]
Clustering Instances	Including many sample graphs	Alex Arenas[1] DIMACS[2] Pajec[3]	https://www.cc.gatech.edu/dimacs10/archive/clustering.shtml	[172]
Epinions trust	Nodes: Epinion users; Edges: directed links	Epinions.com	https://snap.stanford.edu/data/soc-Epinions1.html	[173]
Slashdot Zoo	Nodes: Slashdot users; Edges: directed links	slashdot.org	https://snap.stanford.edu/data/soc-Slashdot0902.html	[174]
MPI-SWS	Nodes: 54 million users; Edges: 2 billion follow links; Posts: 1.7 billion tweets	Twitter	https://twitter.mpi-sws.org/	[175]
Covid19 conspiracy-theory	Reddit: 4377 English posts 4chan: Posts, Poster ID, Responder ID GELDET: Daily filtering of 50 news reports; 324510 relationships	Reddit / 4chan.org/ Google's GDELT / Conspiracy Theories	https://osf.io/j5tg6/	[112]
MediaEval 2020	For Conspiracy, including English tweets, retweets, replies & quotes related to the COVID-19 Train set: 6 458 tweets; 2, 327 retweet Test set: 3,230 tweets; 1, 165 retweet	Twitter / Conspiracy Theories, Fake News	https://zenodo.org/record/4592249#.Y6sHx-zMJAc	[176]

1 http://deim.urv.cat/~aarenas/data/welcome.htm
2 http://www.dis.uniroma1.it/~challenge9/download.shtml
3 http://vlado.fmf.uni-lj.si/pub/networks/data/

REFERENCES

[1] J. Ratkiewicz et al., "Truthy: mapping the spread of astroturf in microblog streams," in Proceedings of the 20th international conference companion on World wide web - WWW '11, Hyderabad, India, 2011, p. 249. doi: 10.1145/1963192.1963301.

[2] N. P. Nguyen, G. Yan, and M. T. Thai, "Analysis of misinformation containment in online social networks," Comput. Netw., vol. 57, no. 10, pp. 2133–2146, Jul. 2013, doi: 10.1016/j.comnet.2013.04.002.

[3] S. Vosoughi, "Automatic Detection and Verification of Rumors on Twitter," 2015.

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[5] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," Nat. Commun., vol. 10, no. 1, p. 7, Jan. 2019, doi: 10.1038/s41467-018-07761-2.

[6] M. S. Islam et al., "COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis," Am. J. Trop. Med. Hyg., vol. 103, no. 4, pp. 1621–1629, Oct. 2020, doi: 10.4269/ajtmh.20-0812.

[7] R. K. Kaliyar and N. Singh, "Misinformation Detection on Online Social Media-A Survey," in 2019 10th International Conference on Computing, Communication and Networking Technologies (icccnt), New York, 2019. Accessed: Feb. 11, 2022. [Online]. Available: http://www.webofscience.com/wos/woscc/full-record/WOS:000525828100164

[8] P. Hernon, "Disinformation and misinformation through the internet: Findings of an exploratory study," Gov. Inf. Q., vol. 12, no. 2, pp. 133–139, Jan. 1995, doi: 10.1016/0740-624X(95)90052-7.

[9] P. N. Petratos, "Misinformation, disinformation, and fake news: Cyber risks to business," Bus. Horiz., Aug. 2021, doi: 10.1016/j.bushor.2021.07.012.

[10] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating Fake News: A Survey on Identification and Mitigation Techniques," ArXiv190106437 Cs Stat, Jan. 2019, Accessed: Feb. 15, 2022. [Online]. Available: http://arxiv.org/abs/1901.06437

[11] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and Resolution of Rumours in Social Media: A Survey," ACM Comput. Surv., vol. 51, no. 2, p. 32:1-32:36, Feb. 2018, doi: 10.1145/3161603.

[12] R. Sicilia, S. Lo Giudice, Y. Pei, M. Pechenizkiy, and P. Soda, "Twitter rumour detection in the health domain," Expert Syst. Appl., vol. 110, pp. 33–40, Nov. 2018, doi: 10.1016/j.eswa.2018.05.019.

[13] G. Tong, W. Wu, and D.-Z. Du, "Distributed Rumor Blocking With Multiple Positive Cascades," IEEE Trans. Comput. Soc. Syst., vol. 5, no. 2, pp. 468–480, Jun. 2018, doi: 10.1109/TCSS.2018.2818661.

[14] A. Saxena, W. Hsu, M. L. Lee, H. Leong Chieu, L. Ng, and L. N. Teow, "Mitigating Misinformation in Online Social Network with Top-k Debunkers and Evolving User Opinions," in Companion Proceedings of the Web Conference 2020, Taipei Taiwan, Apr. 2020, pp. 363–370. doi: 10.1145/3366424.3383297.

[15] J. Jing, F. Li, B. Song, Z. Zhang, and K.-K. R. Choo, "Disinformation Propagation Trend Analysis and Identification Based on Social Situation Analytics and Multilevel Attention Network," IEEE Trans. Comput. Soc. Syst., pp. 1–16, 2022, doi: 10.1109/TCSS.2022.3169132.

[16] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment Analysis for Fake News Detection," Electronics, vol. 10, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/electronics10111348.

[17] D. de Beer and M. Matthee, "Approaches to Identify Fake News: A Systematic Literature Review," in Integrated Science in Digital Age 2020, Cham, 2021, pp. 13–22. doi: 10.1007/978-3-030-49264-9_2.

[18] E. Kapantai, A. Christopoulou, C. Berberidis, and V. Peristeras, "A systematic literature review on disinformation: Toward a unified taxonomical framework," New Media Soc., vol. 23, no. 5, pp. 1301–1326, May 2021, doi: 10.1177/1461444820959296.

[19] Z. I. Mahid, S. Manickam, and S. Karuppayah, "Fake News on Social Media: Brief Review on Detection Techniques," in 2018 Fourth International Conference on Advances in Computing, Communication Automation (ICACCA), Oct. 2018, pp. 1–5. doi: 10.1109/ICACCAF.2018.8776689.

[20] S. I. Manzoor, J. Singla, and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Apr. 2019, pp. 230–234. doi: 10.1109/ICOEI.2019.8862770.

[21] C. T. Mesquita, A. Oliveira, F. L. Seixas, and A. Paes, "Infodemia, Fake News and Medicine: Science and The Quest for Truth," Int. J. Cardiovasc. Sci., vol. 33, pp. 203–205, Apr. 2020, doi: 10.36660/ijcs.20200073.

[22] P. Nakov and G. Da San Martino, "Fake News, Disinformation, Propaganda, Media Bias, and Flattening the Curve of the COVID-19 Infodemic," in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, New York, NY, USA, Aug. 2021, pp. 4054–4055. doi: 10.1145/3447548.3470790.

[23] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating Deep Learning Approaches for Covid19 Fake News Detection," in Combating Online Hostile Posts in Regional Languages during Emergency Situation, Cham, 2021, pp. 153–163. doi: 10.1007/978-3-030-73696-5_15.

[24] L. Bode and E. K. Vraga, "In Related News, That was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media," J. Commun., vol. 65, no. 4, pp. 619–638, Aug. 2015, doi: 10.1111/jcom.12166.

[25] B. C. Stahl, "On the Difference or Equality of Information, Misinformation, and Disinformation: A Critical Research Perspective," Informing Sci. Int. J. Emerg. Transdiscipl., vol. 9, pp. 083–096, 2006.

[26] V. L. Rubin, Y. Chen, and N. K. Conroy, "Deception detection for news: Three types of fakes," Proc. Assoc. Inf. Sci. Technol., vol. 52, no. 1, pp. 1–4, 2015, doi: 10.1002/pra2.2015.145052010083.

[27] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, Oct. 2020, doi: 10.1145/3395046.

[28] M. Celliers and M. Hattingh, "A Systematic Review on Fake News Themes Reported in Literature," in Responsible Design, Implementation and Use of Information and Communication Technology, Cham, 2020, pp. 223–234. doi: 10.1007/978-3-030-45002-1_19.

[29] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manag., vol. 57, no. 2, p. 102025, Mar. 2020, doi: 10.1016/j.ipm.2019.03.004.

[30] H. Allcott, M. Gentzkow, and C. Yu, "Trends in the diffusion of misinformation on social media," Res. Polit., vol. 6, no. 2, p. 2053168019848554, Apr. 2019, doi: 10.1177/2053168019848554.

[31] B. Kim, A. Xiong, D. Lee, and K. Han, "A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions," PLOS ONE, vol. 16, no. 12, p. e0260080, Dec. 2021, doi: 10.1371/journal.pone.0260080.

[32] C. Melchior and M. Oliveira, "Health-related fake news on social media platforms: A systematic literature review," New Media Soc., p. 14614448211038762, Aug. 2021, doi: 10.1177/14614448211038762.

[33] E. C. Tandoc, Z. W. Lim, and R. Ling, "Defining 'Fake News': A typology of scholarly definitions," Digit. Journal., vol. 6, no. 2, pp. 137–153, Feb. 2018, doi: 10.1080/21670811.2017.1360143.

[34] R. Greifeneder, M. E. Jaffé, E. J. Newman, and N. Schwarz, Eds., The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation. London: Routledge, 2020. doi: 10.4324/9780429295379.

[35] S. Kumar and N. Shah, "False Information on Web and Social Media: A Survey," Apr. 2018, doi: 10.48550/arXiv.1804.08559.

[36] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," Big Data, vol. 8, no. 3, pp. 171–188, Jun. 2020, doi: 10.1089/big.2020.0062.

[37] M. T. says, "Fake news. It's complicated.," First Draft, Feb. 16, 2017. https://firstdraftnews.org:443/articles/fake-news-complicated/ (accessed Feb. 19, 2022).

[38] D. Fallis, "A Conceptual Analysis of Disinformation," Feb. 2009, Accessed: Sep. 17, 2021. [Online]. Available: https://www.ideals.illinois.edu/handle/2142/15205

[39] B. Guo, Y. Ding, L. Yao, Y. Liang, and Z. Yu, "The Future of False Information Detection on Social Media: New Perspectives and Trends," ACM Comput. Surv., vol. 53, no. 4, p. 68:1-68:36, Jul. 2020, doi: 10.1145/3393880.

[40] G. D. Domenico, J. Sit, A. Ishizaka, and D. Nunan, “Fake news, social media and marketing: A systematic review,” J. Bus. Res., vol. 124, pp. 329–341, Jan. 2021, doi: 10.1016/j.jbusres.2020.11.037.
[41] P. Meel and D. K. Vishwakarma, “Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities,” Expert Syst. Appl., vol. 153, p. 112986, Sep. 2020, doi: 10.1016/j.eswa.2019.112986.
[42] H. Allcott and M. Gentzkow, “Social Media and Fake News in the 2016 Election,” J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, May 2017, doi: 10.1257/jep.31.2.211.
[43] J. Brummette, M. DiStaso, M. Vafeiadis, and M. Messner, “Read All About It: The Politicization of ‘Fake News’ on Twitter,” Journal. Mass Commun. Q., vol. 95, no. 2, pp. 497–517, Jun. 2018, doi: 10.1177/1077699018769906.
[44] D. Klein and J. Wueller, “Fake News: A Legal Perspective,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2958790, Mar. 2017. Accessed: Mar. 01, 2022. [Online]. Available: https://papers.ssrn.com/abstract=2958790
[45] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, “Rumor has it: Identifying Misinformation in Microblogs,” in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK., Jul. 2011, pp. 1589–1599. Accessed: Mar. 01, 2022. [Online]. Available: https://aclanthology.org/D11-1147
[46] N. DiFonzo and P. Bordia, “Rumor, Gossip and Urban Legends,” Diogenes, vol. 54, no. 1, pp. 19–35, Feb. 2007, doi: 10.1177/0392192107073433.
[47] S. A. Alkhodair, S. H. H. Ding, B. C. M. Fung, and J. Liu, “Detecting breaking news rumors of emerging topics in social media,” Inf. Process. Manag., vol. 57, no. 2, p. 102018, Mar. 2020, doi: 10.1016/j.ipm.2019.02.016.
[48] J. Golbeck et al., “Fake News vs Satire: A Dataset and Analysis,” in Proceedings of the 10th ACM Conference on Web Science, New York, NY, USA, May 2018, pp. 17–21. doi: 10.1145/3201064.3201100.
[49] C. Pigden, “Popper Revisited, or What Is Wrong With Conspiracy Theories?,” Philos. Soc. Sci., vol. 25, no. 1, pp. 3–34, Mar. 1995, doi: 10.1177/004839319502500101.
[50] B. L. Keeley, “Of Conspiracy Theories,” J. Philos., vol. 96, no. 3, pp. 109–126, 1999, doi: 10.2307/2564659.
[51] S. Clarke, “Conspiracy Theories and Conspiracy Theorizing,” Philos. Soc. Sci., vol. 32, no. 2, pp. 131–150, Jun. 2002, doi: 10.1177/004931032002001.
[52] A. M. Enders et al., “The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation,” Polit. Behav., Jul. 2021, doi: 10.1007/s11109-021-09734-6.
[53] C. Shao et al., “Anatomy of an online misinformation network,” PLOS ONE, vol. 13, no. 4, p. e0196087, Apr. 2018, doi: 10.1371/journal.pone.0196087.
[54] G. L. Ciampaglia, A. Mantzarlis, G. Maus, and F. Menczer, “Research Challenges of Digital Misinformation: Toward a Trustworthy Web,” AI Mag., vol. 39, no. 1, Art. no. 1, Mar. 2018, doi: 10.1609/aimag.v39i1.2783.
[55] D. Lazer et al., “Combating fake news: an agenda for research and action,” Report, May 2017. Accessed: Apr. 18, 2022. [Online]. Available: https://apo.org.au/node/76233
[56] M. McClure Haughey, M. D. Muralikumar, C. A. Wood, and K. Starbird, “On the Misinformation Beat: Understanding the Work of Investigative Journalists Reporting on Problematic Information Online,” Proc. ACM Hum.-Comput. Interact., vol. 4, no. CSCW2, pp. 1–22, Oct. 2020, doi: 10.1145/3415204.
[57] M. Fernandez and H. Alani, “Online Misinformation: Challenges and Future Directions,” in Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW ’18, Lyon, France, 2018, pp. 595–602. doi: 10.1145/3184558.3188730.
[58] C. Silverman, A. Hooper, and J. Jurich, “Emergent,” 2015. http://www.emergent.info/ (accessed Apr. 18, 2022).
[59] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, “TweetCred: Real-Time Credibility Assessment of Content on Twitter,” ArXiv14055490 Phys., Jan. 2015, Accessed: Apr. 18, 2022. [Online]. Available: http://arxiv.org/abs/1405.5490
[60] P. T. Metaxas, S. Finn, and E. Mustafaraj, “Using TwitterTrails.com to Investigate Rumor Propagation,” in Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, New York, NY, USA, Feb. 2015, pp. 69–72. doi: 10.1145/2685553.2702691.
[61] T. Mitra and E. Gilbert, “CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations,” in ICWSM, 2015.
[62] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, “Hoaxy: A Platform for Tracking Online Misinformation,” Proc. 25th Int. Conf. Companion World Wide Web - WWW 16 Companion, pp. 745–750, 2016, doi: 10.1145/2872518.2890098.
[63] J. Cao, J. Guo, X. Li, Z. Jin, H. Guo, and J. Li, “Automatic Rumor Detection on Microblogs: A Survey,” ArXiv180703505 Cs, Jul. 2018, Accessed: Aug. 26, 2021. [Online]. Available: http://arxiv.org/abs/1807.03505
[64] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, “Trends in combating fake news on social media – a survey,” J. Inf. Telecommun., vol. 5, no. 2, pp. 247–266, Apr. 2021, doi: 10.1080/24751839.2020.1847379.
[65] M. R. Islam, S. Liu, X. Wang, and G. Xu, “Deep learning for misinformation detection on online social networks: a survey and new perspectives,” Soc. Netw. Anal. Min., vol. 10, no. 1, p. 82, Sep. 2020, doi: 10.1007/s13278-020-00696-x.
[66] D. Varshney and D. K. Vishwakarma, “A review on rumour prediction and veracity assessment in online social network,” Expert Syst. Appl., vol. 168, p. 114208, Apr. 2021, doi: 10.1016/j.eswa.2020.114208.
[67] B. Guo, Y. Ding, L. Yao, Y. Liang, and Z. Yu, “The Future of Misinformation Detection: New Perspectives and Trends.” arXiv, Sep. 09, 2019. Accessed: Nov. 29, 2022. [Online]. Available: http://arxiv.org/abs/1909.03654
[68] S. Shelke and V. Attar, “Source detection of rumor in social network – A review,” Online Soc. Netw. Media, vol. 9, pp. 30–42, Jan. 2019, doi: 10.1016/j.osnem.2018.12.001.
[69] A. Bondielli and F. Marcelloni, “A survey on fake news and rumour detection techniques,” Inf. Sci., vol. 497, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05.035.
[70] M. Almaliki, “Online Misinformation Spread: A Systematic Literature Map,” in Proceedings of the 2019 3rd International Conference on Information System and Data Mining, New York, NY, USA, Apr. 2019, pp. 171–178. doi: 10.1145/3325917.3325938.
[71] K. P. K. Kumar and G. Geethakumari, “Detecting misinformation in online social networks using cognitive psychology,” Hum.-Centric Comput. Inf. Sci., vol. 4, no. 1, p. 14, Sep. 2014, doi: 10.1186/s13673-014-0014-x.
[72] X. Yao, Y. Gu, C. Gu, and H. Huang, “Fast controlling of rumors with limited cost in social networks,” Comput. Commun., vol. 182, pp. 41–51, Jan. 2022, doi: 10.1016/j.comcom.2021.10.041.
[73] M. Cheng, S. Nazarian, and P. Bogdan, “VRoC: Variational Autoencoder-aided Multi-task Rumor Classifier Based on Text,” in Proceedings of The Web Conference 2020, Taipei Taiwan, Apr. 2020, pp. 2892–2898. doi: 10.1145/3366423.3380054.
[74] D. Shah and T. Zaman, “Rumors in a Network: Who’s the Culprit?,” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5163–5181, Aug. 2011, doi: 10.1109/TIT.2011.2158885.
[75] M. Amoruso, D. Anello, V. Auletta, R. Cerulli, D. Ferraioli, and A. Raiconi, “Contrasting the Spread of Misinformation in Online Social Networks,” J. Artif. Intell. Res., vol. 69, pp. 847–879, Nov. 2020, doi: 10.1613/jair.1.11509.
[76] P. Watine, A. Bodaghi, and K. A. Schmitt, “Can the Hawkes process be used to evaluate the spread of online information?,” in 2021 IEEE International Symposium on Technology and Society (ISTAS), Oct. 2021, pp. 1–6. doi: 10.1109/ISTAS52410.2021.9629133.
[77] Z. Shae and J. Tsai, “AI Blockchain Platform for Trusting News,” in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, pp. 1610–1619. doi: 10.1109/ICDCS.2019.00160.
[78] V. L. Rubin and T. Lukoianova, “Truth and deception at the rhetorical structure level: Truth and Deception at the Rhetorical Structure Level,” J. Assoc. Inf. Sci. Technol., vol. 66, no. 5, pp. 905–917, May 2015, doi: 10.1002/asi.23216.
[79] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan, “A Convolutional Approach for Misinformation Identification,” pp. 3901–3907, 2017.
[80] J. L. Elman, “Finding structure in time,” Cogn. Sci., vol. 14, no. 2, pp. 179–211, Apr. 1990, doi: 10.1016/0364-0213(90)90002-E.
[81] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[82] M. Cheng, Y. Li, S. Nazarian, and P. Bogdan, “From rumor to genetic mutation detection with explanations: a GAN approach,” Sci. Rep., vol. 11, no. 1, p. 5861, Mar. 2021, doi: 10.1038/s41598-021-84993-1.
[83] I. Goodfellow et al., “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, 2014, vol. 27. Accessed: Mar. 09, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html
[84] R. S. Sutton and A. G. Barto, “Reinforcement learning: an introduction, ser. Adaptive computation and machine learning,” Camb. Mass. MIT Press, vol. 6, pp. 15–17, 1998.
[85] J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich, “NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube,” in Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, Online, Jul. 2020. Accessed: Jan. 21, 2022. [Online]. Available: https://aclanthology.org/2020.nlpcovid19-acl.17
[86] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv, arXiv:1810.04805, May 2019. doi: 10.48550/arXiv.1810.04805.
[87] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv, arXiv:1907.11692, Jul. 2019. doi: 10.48550/arXiv.1907.11692.
[88] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs,” in Proceedings of the 25th ACM international conference on Multimedia, New York, NY, USA, Oct. 2017, pp. 795–816. doi: 10.1145/3123266.3123454.
[89] S. A. Alkhodair, B. C. M. Fung, S. H. H. Ding, W. K. Cheung, and S.-C. Huang, “Detecting High-Engaging Breaking News Rumors in Social Media,” ACM Trans. Manag. Inf. Syst., vol. 12, no. 1, p. 8:1-8:16, Dec. 2021, doi: 10.1145/3416703.
[90] J. F. Low, B. C. M. Fung, F. Iqbal, and S.-C. Huang, “Distinguishing between fake news and satire with transformers,” Expert Syst. Appl., vol. 187, p. 115824, Jan. 2022, doi: 10.1016/j.eswa.2021.115824.
[91] B. D. Horne and S. Adali, “This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News,” ArXiv170309398 Cs, Mar. 2017, Accessed: Nov. 25, 2021. [Online]. Available: http://arxiv.org/abs/1703.09398
[92] L. Li, G. Cai, and N. Chen, “A Rumor Events Detection Method Based on Deep Bidirectional GRU Neural Network,” in 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, Jun. 2018, pp. 755–759. doi: 10.1109/ICIVC.2018.8492819.
[93] G. S. Cheema, S. Hakimov, E. Müller-Budack, and R. Ewerth, “On the Role of Images for Analyzing Claims in Social Media.” arXiv, Mar. 17, 2021. Accessed: Dec. 26, 2022. [Online]. Available: http://arxiv.org/abs/2103.09602
[94] T. Chen, X. Li, H. Yin, and J. Zhang, “Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection,” in Trends and Applications in Knowledge Discovery and Data Mining, vol. 11154, M. Ganji, L. Rashidi, B. C. M. Fung, and C. Wang, Eds. Cham: Springer International Publishing, 2018, pp. 40–52. doi: 10.1007/978-3-030-04503-6_4.
[95] J. MA et al., “Detecting rumors from microblogs with recurrent neural networks,” Proc. 25th Int. Jt. Conf. Artif. Intell. IJCAI 2016, pp. 3818–3824, Jul. 2016.
[96] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[97] X. Chen, F. Zhou, F. Zhang, and M. Bonsangue, “Catch me if you can: A participant-level rumor detection framework via fine-grained user representation learning,” Inf. Process. Manag., vol. 58, no. 5, p. 102678, Sep. 2021, doi: 10.1016/j.ipm.2021.102678.
[98] T. T. Nguyen, T. T. Nguyen, T. T. Nguyen, B. Vo, J. Jo, and Q. V. H. Nguyen, “JUDO: Just-in-time rumour detection in streaming social platforms,” Inf. Sci., vol. 570, pp. 70–93, Sep. 2021, doi: 10.1016/j.ins.2021.04.018.
[99] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, “Prominent Features of Rumor Propagation in Online Social Media,” in 2013 IEEE 13th International Conference on Data Mining, Dec. 2013, pp. 1103–1108. doi: 10.1109/ICDM.2013.61.
[100] K. Wu, S. Yang, and K. Q. Zhu, “False rumors detection on Sina Weibo by propagation structures,” 2015 IEEE 31st Int. Conf. Data Eng., 2015, doi: 10.1109/ICDE.2015.7113322.
[101] J. Ma, W. Gao, and K.-F. Wong, “Rumor Detection on Twitter with Tree-structured Recursive Neural Networks,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, Jul. 2018, pp. 1980–1989. doi: 10.18653/v1/P18-1184.
[102] S. Lathiya, J. S. Dhobi, A. Zubiaga, M. Liakata, and R. Procter, “Birds of a feather check together: Leveraging homophily for sequential rumour detection,” Online Soc. Netw. Media, vol. 19, 2020, doi: 10.1016/j.osnem.2020.100097.
[103] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[104] S. R. Sangwan and M. Bhatia, “Denigration Bullying Resolution using Wolf Search Optimized Online Reputation Rumour Detection,” Procedia Comput. Sci., vol. 173, pp. 305–314, Jan. 2020, doi: 10.1016/j.procs.2020.06.036.
[105] P. Zhang, H. Ran, C. Jia, X. Li, and X. Han, “A lightweight propagation path aggregating network with neural topic model for rumor detection,” Neurocomputing, vol. 458, pp. 468–477, Oct. 2021, doi: 10.1016/j.neucom.2021.06.062.
[106] T. Ma, H. Zhou, Y. Tian, and N. Al-Nabhan, “A novel rumor detection algorithm based on entity recognition, sentence reconfiguration, and ordinary differential equation network,” Neurocomputing, vol. 447, pp. 224–234, 2021, doi: 10.1016/j.neucom.2021.03.055.
[107] K. Tu, C. Chen, C. Hou, J. Yuan, J. Li, and X. Yuan, “Rumor2vec: A rumor detection framework with joint text and propagation structure representation learning,” Inf. Sci., vol. 560, pp. 137–151, Jun. 2021, doi: 10.1016/j.ins.2020.12.080.
[108] Y. Luo, J. Ma, and C. K. Yeo, “BCMM: A novel post-based augmentation representation for early rumour detection on social media,” Pattern Recognit., vol. 113, p. 107818, May 2021, doi: 10.1016/j.patcog.2021.107818.
[109] S. Sharma and R. Sharma, “Identifying Possible Rumor Spreaders on Twitter: A Weak Supervised Learning Approach,” in 2021 International Joint Conference on Neural Networks (IJCNN), Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9534185.
[110] Z. Wei, X. Xiao, G. Hu, B. Zhang, Q. Li, and S. Xia, “A Novel and High-Accuracy Rumor Detection Approach using Kernel Subtree and Deep Learning Networks,” in 2021 International Joint Conference on Neural Networks (IJCNN), Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9534311.
[111] A. Lao, C. Shi, and Y. Yang, “Rumor Detection with Field of Linear and Non-Linear Propagation,” in Proceedings of the Web Conference 2021, New York, NY, USA, Jun. 2021, pp. 3178–3187. doi: 10.1145/3442381.3450016.
[112] S. Shahsavari, P. Holur, T. Wang, T. R. Tangherlini, and V. Roychowdhury, “Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news,” J. Comput. Soc. Sci., vol. 3, no. 2, pp. 279–317, Nov. 2020, doi: 10.1007/s42001-020-00086-5.
[113] A. Zubiaga, M. Liakata, and R. Procter, “Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media.” arXiv, Oct. 24, 2016. Accessed: Dec. 02, 2022. [Online]. Available: http://arxiv.org/abs/1610.07363
[114] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, “Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vancouver, Canada, Jul. 2017, pp. 647–653. doi: 10.18653/v1/P17-2102.
[115] L. M. S. Khoo, H. L. Chieu, Z. Qian, and J. Jiang, “Interpretable Rumor Detection in Microblogs by Attending to User Interactions,” Proc. AAAI Conf. Artif. Intell., vol. 34, no. 05, Art. no. 05, Apr. 2020, doi: 10.1609/aaai.v34i05.6405.
[116] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in Proceedings of the 20th international conference on World wide web - WWW ’11, Hyderabad, India, 2011, p. 675. doi: 10.1145/1963405.1963500.
[117] N. Hassan, C. Li, and M. Tremayne, “Detecting Check-worthy Factual Claims in Presidential Debates,” in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne Australia, Oct. 2015, pp. 1835–1838. doi: 10.1145/2806416.2806652.
[118] S. Carton, S. Park, N. Zeffer, E. Adar, Q. Mei, and P. Resnick, “Audience Analysis for Competing Memes in Social Media,” Proc.

Int. AAAI Conf. Web Soc. Media, vol. 9, no. 1, Art. no. 1, 2015, doi: 10.1609/icwsm.v9i1.14632.

[119] S. Cohen, J. T. Hamilton, and F. Turner, "Computational Journalism," 2011. https://cacm.acm.org/magazines/2011/10/131400-computational-journalism/abstract (accessed Nov. 25, 2022).

[120] G. Li et al., "Deep reinforcement learning based ensemble model for rumor tracking," Inf. Syst., vol. 103, p. 101772, Jan. 2022, doi: 10.1016/j.is.2021.101772.

[121] M. Lukasik, P. K. Srijith, D. Vu, K. Bontcheva, A. Zubiaga, and T. Cohn, "Hawkes Processes for Continuous Time Sequence Classification: an Application to Rumour Stance Classification in Twitter," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, Aug. 2016, pp. 393–398. doi: 10.18653/v1/P16-2064.

[122] Z. Jahanbakhsh-Nagadeh, M.-R. Feizi-Derakhshi, and A. Sharifi, "A semi-supervised model for Persian rumor verification based on content information," Multimed. Tools Appl., vol. 80, no. 28, pp. 35267–35295, Nov. 2021, doi: 10.1007/s11042-020-10077-3.

[123] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and Resolution of Rumours in Social Media: A Survey," ACM Comput. Surv., vol. 51, no. 2, pp. 1–36, Mar. 2019, doi: 10.1145/3161603.

[124] S. Hamidian and M. T. Diab, "Rumor Detection and Classification for Twitter Data," ArXiv, 2019.

[125] K. Jaidka, K. Ramachandran, P. Gupta, and S. Rustagi, "SocialStories: Segmenting Stories within Trending Twitter Topics," in Proceedings of the 3rd IKDD Conference on Data Science, 2016, Pune India, Mar. 2016, pp. 1–7. doi: 10.1145/2888451.2888453.

[126] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: can we trust what we RT?," in Proceedings of the First Workshop on Social Media Analytics - SOMA '10, Washington D.C., District of Columbia, 2010, pp. 71–79. doi: 10.1145/1964858.1964869.

[127] R. Procter, F. Vis, and A. Voss, "Reading the riots on Twitter: methodological innovation for the analysis of big data," Int. J. Soc. Res. Methodol., vol. 16, no. 3, pp. 197–214, May 2013, doi: 10.1080/13645579.2013.774172.

[128] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 Task 6: Detecting Stance in Tweets," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, California, Jun. 2016, pp. 31–41. doi: 10.18653/v1/S16-1003.

[129] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts," in Proceedings of the 24th International Conference on World Wide Web, Republic and Canton of Geneva, CHE, May 2015, pp. 1395–1405. doi: 10.1145/2736277.2741637.

[130] X. Liu et al., "Reuters Tracer: A Large Scale System of Detecting & Verifying Real-Time News Events from Twitter," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, New York, NY, USA, Oct. 2016, pp. 207–216. doi: 10.1145/2983323.2983363.

[131] E. Kochkina, M. Liakata, and A. Zubiaga, "All-in-one: Multi-task Learning for Rumour Verification," ArXiv180603713 Cs, Jun. 2018, Accessed: Aug. 26, 2021. [Online]. Available: http://arxiv.org/abs/1806.03713

[132] J. Ma, W. Gao, and K.-F. Wong, "Detect Rumor and Stance Jointly by Neural Multi-task Learning," in Companion Proceedings of the The Web Conference 2018, Republic and Canton of Geneva, CHE, Apr. 2018, pp. 585–593. doi: 10.1145/3184558.3188729.

[133] Q. Li, Q. Zhang, and L. Si, "Rumor Detection by Exploiting User Credibility Information, Attention and Multi-task Learning," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, Jul. 2019, pp. 1173–1179. doi: 10.18653/v1/P19-1113.

[134] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[135] X. Wang, Y. Li, J. Li, Y. Liu, and C. Qiu, "A rumor reversal model of online health information during the Covid-19 epidemic," Inf. Process. Manag., vol. 58, no. 6, p. 102731, Nov. 2021, doi: 10.1016/j.ipm.2021.102731.

[136] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez-Rodriguez, "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, New York, NY, USA, 2018, pp. 324–332. doi: 10.1145/3159652.3159734.

[137] N. Turenne, "The rumour spectrum," PLOS ONE, vol. 13, no. 1, p. e0189080, Jan. 2018, doi: 10.1371/journal.pone.0189080.

[138] S. Lomborg and A. Bechmann, "Using APIs for Data Collection on Social Media," Inf. Soc., vol. 30, no. 4, pp. 256–265, Aug. 2014, doi: 10.1080/01972243.2014.915276.

[139] X. K. Chen, J.-C. Na, L. K.-W. Tan, M. Chong, and M. Choy, "Exploring how online responses change in response to debunking messages about COVID-19 on WhatsApp," Online Inf. Rev., vol. 46, no. 6, pp. 1184–1204, Jan. 2022, doi: 10.1108/OIR-08-2021-0422.

[140] A. de Paz, M. Suárez, S. Palmero, S. Degli-Esposti, and D. Arroyo, "Following Negationists on Twitter and Telegram: Application of NCD to the Analysis of Multiplatform Misinformation Dynamics," in Proceedings of the International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2022), Cham, 2023, pp. 1110–1116. doi: 10.1007/978-3-031-21333-5_110.

[141] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: a survey of evaluation datasets," PeerJ Comput. Sci., vol. 7, p. e518, Jun. 2021, doi: 10.7717/peerj-cs.518.

[142] A. Zubiaga, M. Liakata, and R. Procter, "Exploiting Context for Rumour Detection in Social Media," in Social Informatics, Cham, 2017, pp. 109–123. doi: 10.1007/978-3-319-67217-5_8.

[143] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. Wong Sak Hoi, and A. Zubiaga, "SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, Canada, Aug. 2017, pp. 69–76. doi: 10.18653/v1/S17-2006.

[144] G. Gorrell, K. Bontcheva, L. Derczynski, E. Kochkina, M. Liakata, and A. Zubiaga, "RumourEval 2019: Determining Rumour Veracity and Support for Rumours," ArXiv180906683 Cs, Sep. 2018, Accessed: Aug. 26, 2021. [Online]. Available: http://arxiv.org/abs/1809.06683

[145] A. Miani, T. Hills, and A. Bangerter, "LOCO: The 88-million-word language of conspiracy corpus," Behav. Res. Methods, vol. 54, no. 4, pp. 1794–1817, Aug. 2022, doi: 10.3758/s13428-021-01698-z.

[146] R. Aly et al., "FEVEROUS: Fact Extraction and VERification Over Unstructured and Structured information," ArXiv210605707 Cs, Oct. 2021, Accessed: Nov. 25, 2021. [Online]. Available: http://arxiv.org/abs/2106.05707

[147] S. Helmstetter and H. Paulheim, "Collecting a Large Scale Dataset for Classifying Fake News Tweets Using Weak Supervision," Future Internet, vol. 13, no. 5, Art. no. 5, May 2021, doi: 10.3390/fi13050114.

[148] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "ArCOV-19: The First Arabic COVID-19 Twitter Dataset with Propagation Networks." arXiv, Mar. 13, 2021. Accessed: Nov. 25, 2022. [Online]. Available: http://arxiv.org/abs/2004.05861

[149] H. Mubarak and S. Hassan, "ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic." arXiv, Mar. 01, 2021. Accessed: Nov. 28, 2022. [Online]. Available: http://arxiv.org/abs/2012.01462

[150] S. Alqurashi, B. Hamoui, A. Alashaikh, A. Alhindi, and E. Alanazi, "Eating Garlic Prevents COVID-19 Infection: Detecting Misinformation on the Arabic Content of Twitter." arXiv, Jan. 09, 2021. Accessed: Nov. 28, 2022. [Online]. Available: http://arxiv.org/abs/2101.05626

[151] J. M. Banda et al., "A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration," Epidemiologia, vol. 2, no. 3, pp. 315–324, Aug. 2021, doi: 10.3390/epidemiologia2030024.

[152] F. Arslan, N. Hassan, C. Li, and M. Tremayne, "A Benchmark Dataset of Check-worthy Factual Claims." arXiv, Apr. 29, 2020. doi: 10.48550/arXiv.2004.14425.

[153] R. Lamsal, "Coronavirus (COVID-19) Tweets Dataset." IEEE, Mar. 13, 2020. Accessed: Nov. 30, 2022. [Online]. Available: https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset

[154] J. Thorne, A. Vlachos, O. Cocarascu, C. Christodoulopoulos, and A. Mittal, "The FEVER2.0 Shared Task," in Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER), Hong Kong, China, Nov. 2019, pp. 1–6. doi: 10.18653/v1/D19-6601.

[155] J. Norregaard, B. D. Horne, and S. Adali, "NELA-GT-2018: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles." arXiv, Apr. 02, 2019. Accessed: Dec. 26, 2022. [Online]. Available: http://arxiv.org/abs/1904.01546

[156] N. T. Tam, M. Weidlich, B. Zheng, H. Yin, N. Q. V. Hung, and B. Stantic, "From anomaly detection to rumour detection using data streams of social platforms," Proc. VLDB Endow., vol. 12, no. 9, pp. 1016–1029, May 2019, doi: 10.14778/3329772.3329778.
[157] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: a large-scale dataset for Fact Extraction and VERification," ArXiv180305355 Cs, Dec. 2018, Accessed: Nov. 25, 2021. [Online]. Available: http://arxiv.org/abs/1803.05355
[158] S. Kwon, M. Cha, and K. Jung, "Rumor Detection over Varying Time Windows," PLOS ONE, vol. 12, no. 1, p. e0168344, Jan. 2017, doi: 10.1371/journal.pone.0168344.
[159] J. Ma, W. Gao, and K.-F. Wong, "Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, Jul. 2017, pp. 708–717. doi: 10.18653/v1/P17-1066.
[160] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, "Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads," PLOS ONE, vol. 11, no. 3, p. e0150989, Mar. 2016, doi: 10.1371/journal.pone.0150989.
[161] W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California, Jun. 2016, pp. 1163–1168. doi: 10.18653/v1/N16-1138.
[162] C. Burfoot and T. Baldwin, "Automatic Satire Detection: Are You Having a Laugh?," in Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore, Aug. 2009, pp. 161–164. Accessed: Nov. 25, 2021. [Online]. Available: https://aclanthology.org/P09-2041
[163] C. Yang, X. Zhou, and R. Zafarani, "CHECKED: Chinese COVID-19 Fake News Dataset." arXiv, Jun. 12, 2021. Accessed: Nov. 25, 2022. [Online]. Available: http://arxiv.org/abs/2010.09029
[164] A. Gruzd and P. Mai, "Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter," Big Data Soc., vol. 7, no. 2, p. 2053951720938405, Jul. 2020, doi: 10.1177/2053951720938405.
[165] L. Zhiquan, "advogato.tar.bz2." figshare, Mar. 03, 2016. doi: 10.6084/m9.figshare.3082042.v1.
[166] K. Lerman, "Digg_2009 social news votes and graph." figshare, Jan. 11, 2016. doi: 10.6084/m9.figshare.2062467.v1.
[167] R. A. Rossi and N. K. Ahmed, "The Network Data Repository with Interactive Graph Analytics and Visualization," Network Data Repository, 2015. https://networkrepository.com/youtube-links.php (accessed Feb. 05, 2022).
[168] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. election: divided they blog," in Proceedings of the 3rd international workshop on Link discovery, New York, NY, USA, Aug. 2005, pp. 36–43. doi: 10.1145/1134271.1134277.
[169] J. Kunegis, "KONECT: the Koblenz network collection," in Proceedings of the 22nd International Conference on World Wide Web, New York, NY, USA, May 2013, pp. 1343–1350. doi: 10.1145/2487788.2488173.
[170] J. McAuley and J. Leskovec, "Learning to Discover Social Circles in Ego Networks," p. 9, 2012.
[171] J. Yang and J. Leskovec, "Defining and Evaluating Network Communities based on Ground-truth," May 2012, Accessed: Feb. 05, 2022. [Online]. Available: https://arxiv.org/abs/1205.6233v3
[172] D. A. Bader, H. Meyerhenke, P. Sanders, and D. Wagner, "Graph Partitioning and Graph Clustering," Graph Partitioning, 2012, [Online]. Available: https://www.cc.gatech.edu/dimacs10/
[173] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed Networks in Social Media," ArXiv10032424 Phys., Mar. 2010, Accessed: Feb. 05, 2022. [Online]. Available: http://arxiv.org/abs/1003.2424
[174] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters." arXiv, Oct. 08, 2008. doi: 10.48550/arXiv.0810.1355.
[175] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," Proc. Int. AAAI Conf. Web Soc. Media, vol. 4, no. 1, Art. no. 1, May 2010, doi: 10.1609/icwsm.v4i1.14033.
[176] K. Pogorelov et al., "FakeNews: Corona Virus and 5G Conspiracy Task at MediaEval 2020," presented at the MediaEval Benchmarking Initiative for Multimedia Evaluation, 2020. Accessed: Dec. 27, 2022. [Online]. Available: https://www.semanticscholar.org/paper/FakeNews%3A-Corona-Virus-and-5G-Conspiracy-Task-at-Pogorelov-Schroeder/07e1f960812214483c9bfa10b9ab13f563e2cff1

Arezo Bodaghi completed her Master's degree in Information Technology Engineering at Tarbiat Modares University in Tehran, Iran in 2016, and is currently pursuing a Ph.D. in Information Systems Engineering at Concordia University in Montreal, Canada. Her research interests are focused on Machine Learning, Deep Learning, Natural Language Processing.

Ketra A. Schmitt received a Bachelor's degree from Duke University in Environmental Science and Policy in 1996. She worked as a health inspector and statistical analyst before starting graduate school at Carnegie Mellon University in 2001, where she received a Master's degree in Statistics in 2005 and Ph.D. in Engineering and Public Policy in 2006. Dr. Schmitt joined Battelle Memorial Institute from 2006 to 2008 where she led an interdisciplinary team to develop an economic impact assessment of terrorism. She joined the Centre for Engineering in Society at Concordia University in 2008 and received tenure in 2014. Dr. Schmitt applies her expertise in engineering and public policy to problems at the intersection of technology, governance and human behavior. She trains graduate and undergraduate students to apply systems modeling techniques to policy issues with significant scientific and social uncertainty. Her research has been funded by NSERC, FRQSC, Natural Resources Canada, Public Safety Canada, Global Affairs Canada and the Department of National Defence. Her work applying systems methods and agent-based models has been used to inform policy for federal agencies in Canada and the US on topics ranging from social media governance, deference, terrorism and environmental protection.

Pierre Watine received the Bachelor's degree in Software Engineering from the University of Concordia, Montreal, Canada, in 2020. He is currently working toward the Master of Applied Sciences degree in information systems security at that same University. His research interests include cyber and cyber-physical security of critical infrastructures and online data surveillance.

Benjamin C. M. Fung (M'09-SM'13) received his Ph.D. degree in Computing Science from Simon Fraser University in Canada in 2007. He is a Canada Research Chair in Data Mining for Cybersecurity, and a Professor at the School of Information Studies, McGill University in Canada. He also serves as Associate Editors for IEEE Transactions of Knowledge and Data Engineering (TKDE) and Elsevier Sustainable Cities and Society (SCS). He has over 150 refereed publications that span the research forums of data mining, privacy protection, cybersecurity, services computing, and building engineering. Prof. Fung is also a licensed Professional Engineer of software engineering in Ontario, Canada.