IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 11, 2022, accepted February 28, 2022, date of publication March 8, 2022, date of current version March 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157724

Are You a Cyborg, Bot or Human?–A Survey on Detecting Fake News Spreaders

WAJIHA SHAHID, YIRAN LI, DAKOTA STAPLES, GULSHAN AMIN, SAQIB HAKAK, AND ALI GHORBANI, (Senior Member, IEEE)

Faculty of Computer Science, Canadian Institute for Cybersecurity, University of New Brunswick, Fredericton, NB E3B 5A3, Canada

Corresponding author: Wajiha Shahid (wajiha.shahid@unb.ca)

This work was supported in part by the Canadian Institute for Cybersecurity (CIC); in part by the Atlantic Opportunity Agency (ACOA); and in part by the Opportunity New Brunswick (ONB), Canada. The work of Ali Ghorbani was supported by the Tier 1 Canada Research Chair.

ABSTRACT One of the major components of Societal Digitalization is Online social networks (OSNs). OSNs can expose people to different popular trends in various aspects of life and alter people's beliefs, behaviors, and decisions and communication. Social bots and malicious users are the significant sources for spreading misinformation on social media and can pose serious cyber threats in society. The degree of similarity of user profiles of a cyber bot and a malicious user spreading fake news is so great that it is very difficult to differentiate both based on their attributes. Over the years, researchers have attempted to find a way to mitigate this problem. However, the detection of fake news spreaders across OSNs remains a challenge. In this paper, we have provided a comprehensive survey of the state of art methods for detecting malicious users and bots based on different features proposed in our novel taxonomy. We have also aimed to avert the crucial problem of fake news detection by discussing several key challenges and potential future research areas to help researchers who are new to this field.

INDEX TERMS Cyborg, deep fake, deceptive content, fake news detection, malicious user, misinformation, news propaganda, social bots, social media.

I. INTRODUCTION

In the present era, our society is gradually getting digitalized as the Internet is the main source of information, entertainment and communication. The central part of societal digitalization is the online social networks(OSNs), such as Facebook, Twitter etc. OSNs are nowadays an integral part of people's daily life, giving users the platform to interact, express themselves and access news [1], [2]. Facebook has 1.88 billion daily users and Twitter has 199 million monetizing daily users [3]. The convenience of social networking has brought the world together due to ease of communication and access to information [4]. At the same time, this easy access to information comes with its drawback such as the excessive propagation of fake news in the form of propaganda, misinformation etc. [5]. More than 40% of traffic to websites spreading fake news is redirected through links on Facebook, Instagram, and Twitter [6] due to their easy access and rapid dissemination [7]. Spread of fake news has even been listed as a major threat to society by the World Economic Forum [8]. Fake news can be described as a kind of news story involving intentional false information to alter users' minds on social media [9]. The dissemination of fake news significantly affects personal reputation and public trust. A survey of 92,000 consumers on a variety of digital topics in 46 markets to see the trust ratio in online news all over the world was conducted by Reuters in 2021[1] as summarized in Figure 1. Results show that Finland had the highest share of respondents agreeing "you can trust news most of the time" at 65% which marks a 9%-point increase since the last edition of the report. The United States made little progress and only 29% of people trusted the news most of the time based on previous experiences.

Although the topic of fake news is not new, the study of fake news spreaders' on social media is a developing topic [10]. There are currently numerous challenging issues [11] which currently require further investigation such

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

[1]https://www.statista.com/chart/7248/where-people-trust-the-news-most-and-least/

as differentiating a user account from automated accounts. Automated accounts are controlled by algorithms known as social bots [12]. Multiple social bots can take the form of a social botnet. Social botnet is a group of social bots created and controlled by a botmaster. They perform malicious activities, such as creating multiple fake accounts, spreading spam, manipulating online ratings, and so on [13].

A recent study[2] estimated that there are 321 million Twitter accounts out of which 48 million are bot accounts, i.e., 15% of all Twitter accounts [14]. The automated nature of bots makes it easy to achieve a large scale impact when spreading misinformation [15]. Analyzing large-scale social data[3] collected during the Catalan referendum for independence on October 1, 2017, consisting of nearly 4 million Twitter posts generated by almost 1 million users revealed that bots produced 23.6% of the total number of posts during the event. A Barracuda report[4] reveals that Automated traffic makes up 64% of internet traffic. Just 25% of it was made up by good bots, while 39% of all traffic was from bad bots as shown in Figure 2. Figure 3 shows the bad bot traffic in North America accounts for 67% of bad bot traffic.

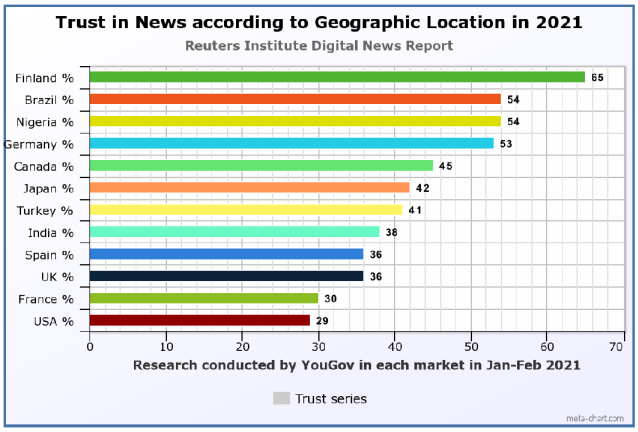


FIGURE 1. Trust in news according to location.

Despite the efforts to detect social bots, it is still difficult to distinguish them from legitimate users which makes it a challenge [16]. The process of social bots identification and their detection is cyclic. New bots are created which spread fake news. And then new social bots filters are derived to tackle them, while old bots mutate into advanced ones [17]. Sometimes automated accounts show human characteristics giving birth to ''Cyborg'' [18]. These bots can even interact as legitimate users when the human takes over the bot profile from time to time.

The aforementioned statistics clearly state the need to come up with an effective solution to identify and detect the fake news spreaders. Various research studies have been

[2]https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[3]https://www.pnas.org/content/115/49/12435

[4]https://www.helpnetsecurity.com/2021/09/07/bad-bots-internet-traffic/

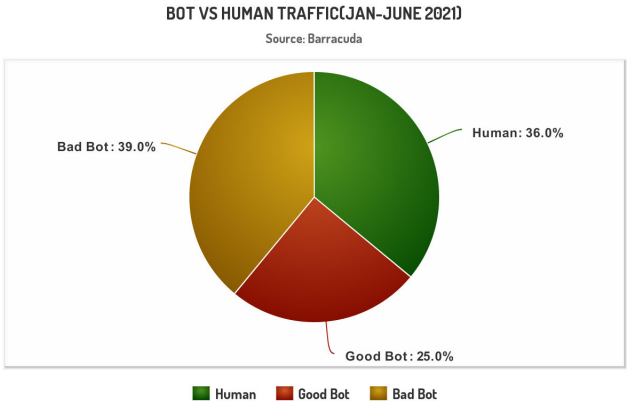


FIGURE 2. Bot Vs human traffic (Jan-June 2021).

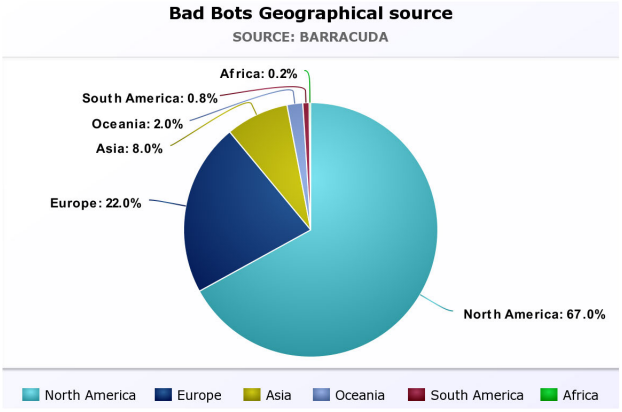


FIGURE 3. Bad bots geographical source.

carried out in the past to identify the nature of the fake news spreaders accounts.

Many surveys have been conducted that reviewed bot detection, human-based detection and cyborg detection along with their taxonomy separately but none of the surveys reviewed all three of them together. The purpose of our survey is to review the recent research done on the subject of bot detection, human detection and cyborg detection. The existing surveys on this topic are summarized in Table 2. Recently in one of the studies, the authors [19] reviewed bot and cyborg detection algorithms, whereas [20] and [21] reviewed only bot detection algorithms, with the former contributing a taxonomy of their work. Authors [22] discussed human user-based detection algorithms [23]. As clearly shown from Table 2, our survey is different from the existing surveys as it not only deals with bot and human-based detection methods but also hybrid-based methods, which include the detection of cyborgs.

To conduct this study, due to an extensive volume of literature on this topic, we used keywords such as Bot detection on social media', 'Fake news users', 'Human and Bot detection on twitter' and other similar keywords. Based on these keywords, relevant papers were extracted published within the last three years from reputed databases such as IEEE,

TABLE 1. Table of acronyms.

Acronym	Explanation
BiGRU	Bidirectional Gated Recurrent Unit
DA SBCD	Deep Auto-encoder based Social Bot Community Detection
DL	Deep Learning
DNN	Deep Neural Network
GRU	Gated Recurrent Unit
LDA	Linear discriminant analysis
LSA	Latent Semantic Analysis
LSTM	Long Short Term Memory
ML	Machine Learning
NLP	Natural Language Processing
NMI	Normalized Mutual Information
OSN	Online Social Network
RGA	Region Growing Algorithm
SBCD	Social Bot Community Detection
SMAN	Structure-aware Multi-head Attention Network
SVM	Support Vector Machine
TF-IDF	Term Frequency Inverse Document

Springer, Elsevier and ACM. From the list of extracted papers, we excluded conference articles, book chapters and shortlisted technical journal articles with a reasonably good number of citations. Furthermore, popular journals on human Psychology were found to search for papers on Fake news targets and characteristics of people who are impacted by fake news. We have also listed the acronyms used in this work in Table 1.

All of our contributions in this paper are summarized as follows:

- An extensive survey over the current state of art methods in detecting bot, human and hybrid-based accounts.
- A novel taxonomy on fake news spreaders detection approaches.
- Identify and discuss existing and emerging new challenges, and future research agendas.

The remaining part of the paper is organized as follows: Section II describes what the fake news is, further discussing its various components and features. In Section III, we propose a taxonomy, and based on that, explain the existing studies. Section IV explains all the methods used to detect the fake news spreaders. In Section V, the challenges and issues in detecting fake news spreaders are discussed. Section VI outlines potential future directions leading to concluding this paper in Section VII.

II. FUNDAMENTALS OF FAKE NEWS

In this section, we discuss the fundamental concepts of fake news. The major fundamental concepts discussed are the definition, components, types, and features of fake news.

A. DEFINITION OF FAKE NEWS

The spread of fake news has become a global issue that needs to be attended immediately [25]. This concept drew major attention after the US elections of 2016 [8]. Fake news is defined by [11], as misleading content including conspiracy theories, rumors, clickbaits, fabricated news, and satire. Reference [26] defines fake news as misinformation and disinformation both, including false and forged information, that is spread on purpose to mislead people or to fulfill a propaganda. In our definition, *"Fake news is a vehicle of purposely targeted fabricated news spread to affect the cognitive activities of a user through user-content interaction by indirectly affecting his unconscious behavior"*. This unconscious behavior can further strengthen confirmation bias among users and aid in further spread of fake news. The purveyors of fake news have been successful as humans have always been attracted to sensationalism and controversies [27]. A recent example is the spread of false information regarding COVID 19 vaccines and dangerous scientific treatment methods posing great risk to public health [28]. Other examples are the political smear campaigns during elections to alter public views about popular candidates and their policies [29]. Figure 4 shows the complete picture of fake news based on its components, features and detection methodologies.

B. COMPONENTS OF FAKE NEWS

In order to clearly understand the spread of fake news, it is important components need to be discussed. These components can be divided into four main categories including creator/spreader, target victims, content and social context [30]. Figure 5 shows how fake news is spread on social media. [31]

1) CREATORS/SPREADERS

Creators generate fake news and spreaders propagate it by re-sharing. They can either be humans or non-humans. Non-humans include social bots and cyborgs. [32]. Social bots are algorithms that are programmed to engage autonomously on social media. They can create content as well as increase its reach [33]. Cyborgs are a hybrid between human accounts and social bots [34].

2) TARGET VICTIMS

Target victims are the group of people or organizations that are impacted by fake news. They are specifically identified and targeted by fake news spreaders. Voters can be targeted in case of smear campaigns during elections [35]. Vulnerable populations include online customers being exposed to scams, patients being exposed to wrong medical information and non-digital natives who don't have enough exposure to differentiate false news from the truth [36].

3) NEWS CONTENT

News content comprises of non-physical and physical contents. Physical contents may include headings and visual features to attract users. Clickbait and hashtags are examples of physical content that catches viewers' attention initially [37]. Non-physical contents contain opinions and sentiments. This is the content that results in creating polarity and change of views. Authors use strong positive or negative emotions to make their content more sensational and easily exploitable [38].

TABLE 2. Previous surveys on the topic vs our survey.

Survey	Year	Contribution of Surveys			
		Bot Detection	Human User-based Detection	Cyborg Detection	Taxonomy
[19]	2019	✓	✗	✓	✗
[23]	2019	✓	✗	✗	✗
[22]	2020	✗	✓	✗	✗
[24]	2021	✓	✗	✗	✓
[20]	2021	✓	✗	✗	✗
[21]	2021	✓	✗	✗	✗
Our Survey	2021	✓	✓	✓	✓

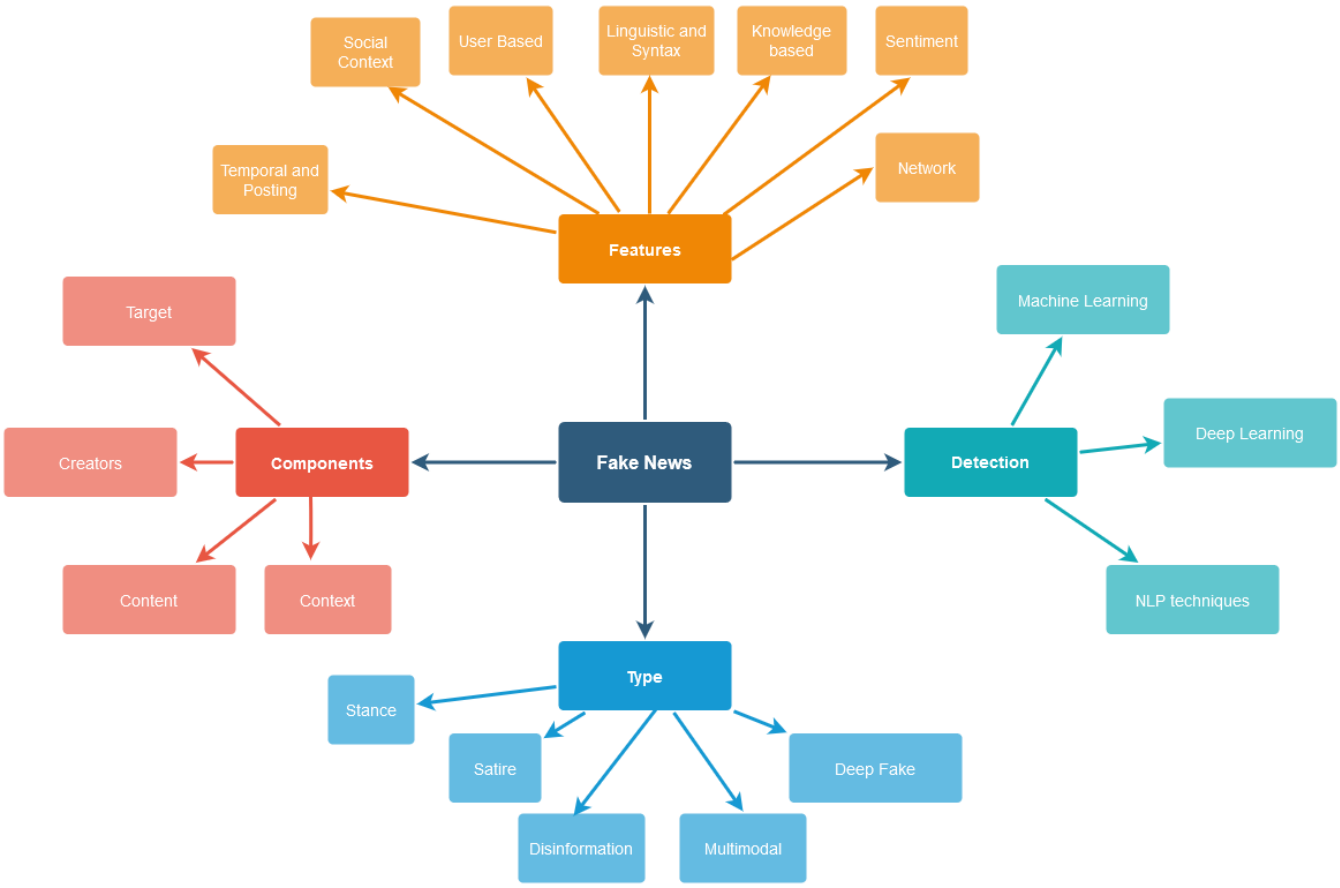


FIGURE 4. Bird eye's view of Fake news.

4) SOCIAL CONTEXT

Social context refers to the overall social environment in which the news is being spread. Social context includes the interaction of online users with each other and the content they are interested in [39]. The social environment of online communities and users, along with the social context, determines how fast fake news is propagated through various channels of social networks [40].

C. TYPES OF FAKE NEWS

Fake news is of multiple kinds. They may be stance, satire, multi-modal, deep fake, and disinformation. *Stance* can be classified into four types, i.e. agree, disagree, discuss and unrelated [42]. An agreeing stance would be related to the headline of fake news, whereas a disagreeing stance holds contradictory information. *Satire* involves humor and mockery [43],it usually includes some political message or criticism in the form of humor and the tone used is generally sarcastic. *Multi-modal* involves spread of fake news using multiple means such as videos,images,audio,text etc. [44]. *Deep fake* is a type of fake news that is spread through manipulated video clips,images and recordings [45]. Deep fake is generated by using deep learning techniques. It lets a computer to generate fabricated media content. A startup named

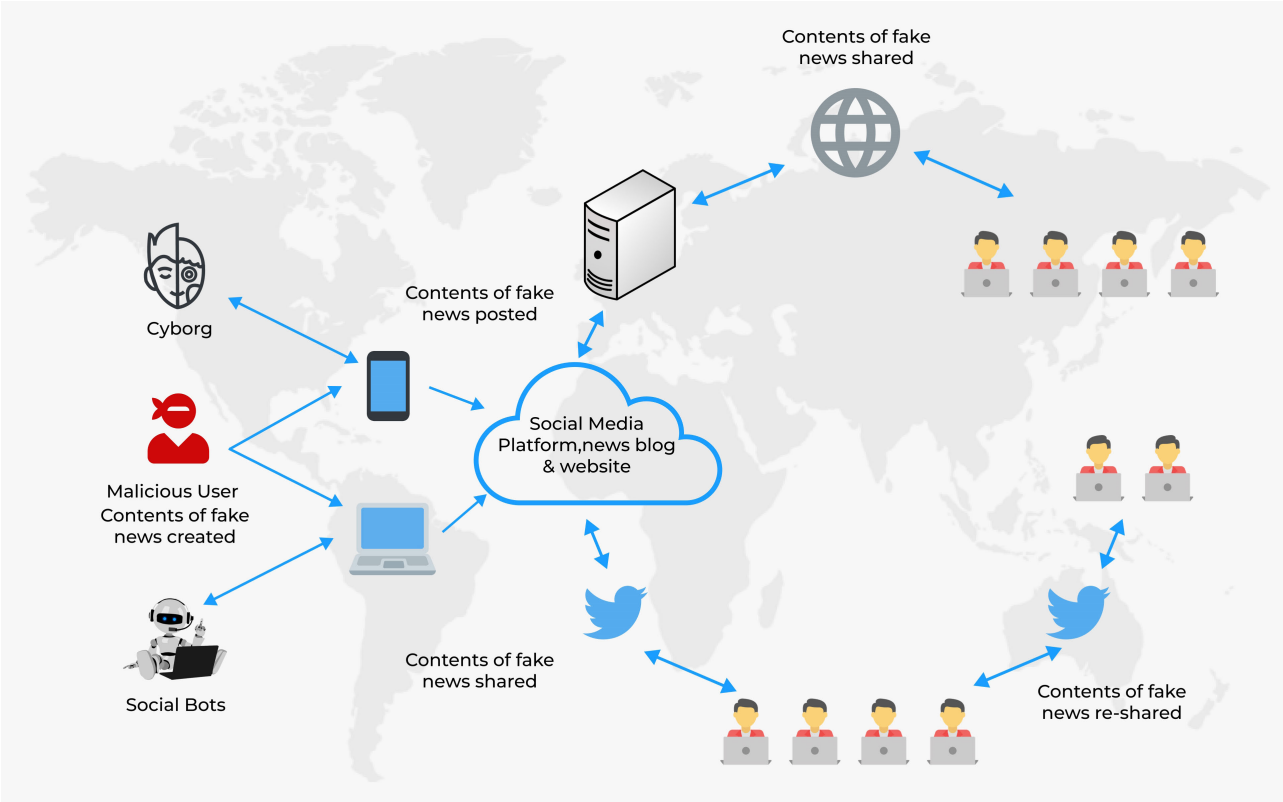


FIGURE 5. Spread of fake news on the social media and internet.

Deep Trace reported 7964 deep fake videos in 2019 and that number doubled within nine months and continues to grow exponentially [46]. *Disinformation* is misleading and false information that is spread in order to deceive people [47]. Disinformation has various sociopolitical repercussions [48]. Sources may spread manipulated information to deceive the audience in order to achieve political agendas or to create social havoc.

D. FEATURES USED IN DETECTING FAKE NEWS

Research work is being carried out in order to help online users uncover and recognize fake news and to develop automatic fake news detection systems. However, this is quite a challenging task as the spreaders of fake news usually come up with newer methods to mislead people [49], [50]. New solutions need to be developed in order to detect fake news sources [51]. In order for the early detection methods to be efficient, features of fake news need to be identified and extracted first [52]. The key features in detecting fake news are user-based features, temporal analysis, sentiment features, linguistic analysis, social context-based analysis, and network features. Details of these features are discussed below:

(i) User-based features include unique characteristics of users profiles that can be analyzed to find out if the person happens to be a fake news spreader [53]. User-based features can be divided into profile analysis and credibility analysis. In *user profile analysis*, user profile can be analyzed on the basis of username, age, profile picture, Geo-location and account verification status [54]. *User credibility analysis* includes information about the number of friends and followers. For example, bot accounts generally have more users in their follower list and follow very less users themselves [55]. *(ii) Temporal analysis* includes timing and frequency of the posts as well as user engagement. This helps to identify bot accounts as they have specific set patterns of online engagement [56]. Bot accounts are programmed to have a more active engagement at certain times. *(iii) Sentiment features* involve analyzing sentiments that trigger emotional response. Bot accounts use out-of-context misleading facts to provoke emotions. A lot of the content created by fake news propagators is highly polarized and exaggerated. Reference [57] *(iv) Linguistic analysis* involves determining the writing patterns and formats [58]. Most fake news creators have a specific writing style and format. Fake news content can be identified by the excessive use of bold letters in headings and paragraphs. The presence of suspicious tokens such as URLs,tags, and excessive uppercase words is also a fake news feature that can be used for detection. *(v) Knowledge-based analysis* involves verifying suspicious information from credible resources such as using verified websites. This can be done manually or by using AI algorithms. *(vi) Social context-based analysis* includes user network analysis and distribution network analyses. User network analysis is used to study the engagement patterns

between online accounts. The distribution pattern focuses on the distribution of information [59]. *(vii) Network features* have two types of networks analyzed: Homogeneous networks and heterogeneous networks. Homogeneous networks have singular nodes and include stance networks and propagation networks [60]. Stance-based modelling determines the users' stance on a specific idea or news. The classification is based on the agreement or disagreement between the main headline and body of the news [61]. Propagation networks analyze the relationship between posts and re-posts. Generally fake news gets re-posted excessively and faster compared to authentic ones [62]. Heterogeneous networks have multiple nodes. It involves analyzing relationships between multiple nodes, including articles, publishers, users and posts [63].

III. FAKE NEWS DISSEMINATION STUDIES BASED ON SPREADERS ACCOUNTS FEATURES

In this section, we will explore state-of-the-art approaches for fake news spreader detection based on our taxonomy as highlighted in Figure 6. All the existing studies are categorized based on three categories, i.e., source, propagation and target based features. We have identified the commonly used datasets in these studies in Table 3.

TABLE 3. Popular datasets used in human and bot detection.

Category	Dataset	Studies
Bot Detection	PyTrawler 2021	[64]
	BadBots	[65]
	Cresci	[66], [67], [68]
	PAN 2020	[69], [70], [71]
	Social Honeypot	[72], [67]
User Detection	Twitter 15	[73], [74], [1]
	Twitter 16	[73], [74], [1]
	FakeNewsNet	[75]
	BuzzFeed	[76], [77]
	PolitiFact	[76], [77], [78], [79], [80]
	Gossipcop	[78]

A. SOURCE-BASED ACCOUNT DETECTION

A source is an originator of fake news [30]. It can either be a human, bot or cyborg [81], [82]. There are different features from which we can identify a source of fake news. We have classified the distinguishing features of source into three main categories i.e., personality feature, historical feature and credibility feature [73]. Table 4 summarizes the studies which have used source-based features to detect fake news using ML, DL and NLP techniques. In the following subsections, the summaries of the existing works along with the feature description are briefly highlighted.

1) PERSONALITY FEATURE

The personality feature includes the qualities of a fake news spreader. It is further divided into the linguistic feature, posting frequency and login interval [1]. *Linguistic features* include the writing style and grammar of the post/tweet [83]. Fake news is generally written in capital letters with typing errors, poor sentence structure, and many exclamation marks. All these come under the linguistic feature of detecting fake news [68]. *Posting frequency* means how many posts or tweets are posted in a day and the time gap between consecutive posts. The *posting frequency* of such accounts is visibly high with repetitive posts being posted after a fixed interval of time [84], [85]. The *login interval* means the time duration of each session and the gap between two consecutive sessions. Fake news spreaders will have longer login intervals, and login time is likely to be the same each day [70], [86].

2) HISTORICAL FEATURE

Historical feature means analyzing the account's metadata and identifying the trends such as login time of the day, posting history and posting time [67]. Metadata provides enough details about the user's profile which helps to identify if it is a real human's account or a bot or cyborg. Similarly, by detecting the *user pattern of logging into his social media account* with respect to time and analyzing the posted tweets, it can detect if the account is a fake news spreader. Spreaders generally spam by posting same fake posts after a fixed interval of time and posting many tweets at a time. *Time-based analysis* can provide a great insight in distinguishing fake news spreader from a normal account [78]. For example, a bot working as fake news spreader will usually post false content all day without break whereas its impossible for a human post all day as humans have other activities too [20].

3) CREDIBILITY FEATURE

The credibility feature takes into account the authenticity of the publishing source and originality of content posted by a user by analyzing the previous posts from that account [87]. Credibility is one of the major features to identify fake news spreaders among other users [76]. Credibility can be assessed on two bases, namely, *publishing source* and *content source*. Posts from unverified or malicious URLs are less authentic and chances are more that they will contain false information or manipulated news [69]. Similarly, the *content source* is also an important feature to distinguish a fake news spreader from other accounts [88]. *Content source* means the platform where information/news is shared. Social platforms like WhatsApp, Facebook, Twitter, Instagram are all accessible to common people and people freely post about literally anything without any justification of their authentication [89]. News posted on these forums is equally likely to be manipulated, if not completely fake [90]. Following studies have used the source features in their detection models.

Lingam *et al.* [72] proposed SBCD and DA-SBCD methods which can detect social bots and identify social botnet communities in online social networks (OSNs). The efficiency of proposed algorithms gives better performance than existing schemes in terms of normalized mutual information (NMI), precision, recall and F-measure. The generated dataset was pre-processed, and machine learning algorithms determined the fake accounts. This paper aimed to identify bots effectively with the minimum possible collection of

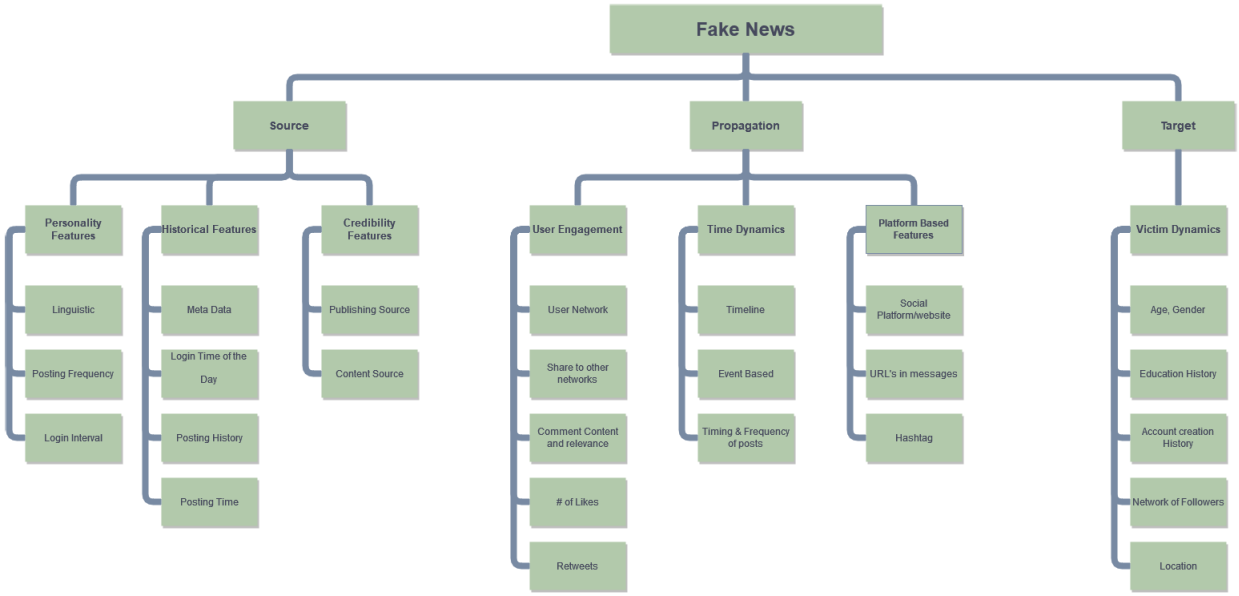


FIGURE 6. Taxonomy on fake news dissemination features based on fake news spreaders accounts.

attributes on the Twitter social network. The authors used the Random Forest classifier to train the data. This study was able to find the percentage of bot accounts in each cluster. For this purpose, a previously trained classifier was used to label the data for bot accounts. This study also concluded that bots had a low follower growth rate as compared to normal accounts. On the other hand, they had high friends/followers ratio. Furthermore, the screen names of bots usually had more digits than a normal account indicating automated behaviors. Apuke and Omar [91] study the underlying reason on who shares fake news, and why they share it. Cardaioli *et al.* [68] used the machine learning approach which could be used in evaluating the stylistic consistency of social network posts and to accomplish other kinds of analyses based on authors style, it can distinguish when posts are posted by cyborgs or bots with statistical evidence. Wu *et al.* [89] used DABot to detect bots and cyborgs by increasing the efficiency of the model by labeling user data and obtaining a large-scale dataset at a small cost. This paper designed a new deep neural network model RGA for detection. Kaliyar *et al.* [76] concluded that not only the content of the news articles is an important factor for fake news detection, but also the existence of echo chambers, a group of users with the same interests grouped together to form a community. Shu *et al.* [78] focus on understanding and exploiting user-profile features on social media for fake news detection. The authors measure users sharing behaviors, and the group representative user sets who are more likely to share news. Khaund *et al.* [20] discuss different methods to detect bots like Early Sybil, Mislove's algorithm and BotGraph. Orabi *et al.* [23] studied the behaviors and features of social media bots to detect the bots who spread fake news online.

B. PROPAGATION-BASED ACCOUNTS DETECTION

A propagator disseminates fake news widely to increase its reach to maximum victims [69]. The features of propagator have been classified into three main categories i.e., user engagement, time dynamics and platform-based features. Table 5 enlists the related existing research mentioning the features of fake news propagator along with the algorithms used.

1) USER ENGAGEMENT FEATURE

The user engagement features include the user network details and circulation of fake news between these networks [84]. Generally fake news propagators have a network of spam accounts [1]. Most of the propagator profiles have a list of spam or bot accounts in their followers and followers list [92]. Furthermore, the content of comments under posts and articles can be analyzed for relevance and to identify fake news propagators [76]. At times, the comments section contain totally irrelevant comments amidst a series of relevant comments [93]. These comments can include malicious URLs or website links advertising something, or even starting a comment war [94]. Most of these irrelevant comments are made to propagate fake news [73]. Moreover, huge number of retweets or re-shares in a small span of time is a good measure to identify fake news propagators as well [88].

2) TIME DYNAMICS FEATURE

The timelines of fake news propagators consist mainly of fake news with a few real news to make them look authentic [89]. There is usually a set pattern following which posts are shared and re-shared on a daily basis [68]. The accounts used by fake news propagators are generally over active and show spam posts and comments at random times. On the other

TABLE 4. Common spreader features.

Study	Source			Techniques		
	Historical	Personality	Credibility	ML	DL	NLP
Lingam et al. (2020) [72]	✗	✗	✓	✗	✓	✗
Apuke et al. (2021) [91]	✗	✗	✓	✓	✗	✗
Cardaioli et al. (2021) [68]	✗	✓	✗	✓	✗	✗
Wu et al. (2021) [89]	✓	✓	✗	✓	✓	✗
Kaliyar et al. (2021) [77]	✗	✓	✓	✓	✓	✗
Khaund et al. (2021) [20]	✓	✓	✗	✓	✓	✗
Orabi et al. (2020) [23]	✓	✓	✓	✓	✗	✗

hand, genuine users usually have specific active and inactive login pattern based on their daily schedule. The fake news propagators are most active during specific important events, such as, election campaigns and social movements [70].

3) PLATFORM-BASED FEATURE

Fake news propagators accounts are found on all popular social media platforms such as Facebook, Instagram, Twitter etc., and even on other community websites like blogs, discussion platforms, fake news websites etc. These accounts are involved in ''hashtag wars'' on social media to propagate fake news and spread propaganda. Moreover, the fake news propagators share irrelevant URLs containing malware links, and unwanted information in comment sections and user inboxes [67]. The following papers discuss the features of fake news propagators.

[95] focused on defining the degree to which bots can exploit hashtags. The machine learning algorithms determined the fake accounts based on preprocessed datasets. This paper aim to identify bots effectively with the minimum possible collection of attributes on the Twitter social network. Authors have used the Random Forest classifier to train the data. This study was able to find the percentage of bot accounts in each cluster. For this purpose, a previously trained classifier was used to label the data for bot accounts. This study also concluded that bots have a low follower growth rate as compared to normal accounts. They have high friends/ followers ratio. Furthermore, the screen names of bots usually have more digital numbers as compared to a normal account indicating automated behavior. Sansonetti *et al.* [80] proposed a text content and social context-based model for human-based fake news detection. Nicola *et al.* [21] proposed a profile feature and timeline based classification of bots, and concluded that feature-based classification deemed to perform well with detecting social and sophisticated bots. Lu and Li [1] proposed a graph-based method to classify a fake news from real one, and highlighted the suspicious re-tweeters. Mendoza *et al.* [94] has proposed a leveraging graph-based representation approach which can learn a network-based representation of users, and is suitable for effectively detecting social bots. This method defines a semi-supervised algorithm which accurately detects groups of social bots, by performing an in-order traversal of the proximity graph. Vogel and Meghana [69] focused on the detection of fake news on Twitter in English and Spanish, and have followed the approach of identifying the fake news spreaders by extracting emotions behind the tweets. Pozzana and Ferrara [84] have analyzed two Twitter datasets: the collection from French presidential election 2017 and the hand-labeled tweets from three groups of bots active in as many viral campaigns, to detect different type of bots. Bello *et al.* [70] proposed a multilingual approach of identifying fake news spreaders on Twitter data. They manually engineered domain-specific features covering behavioural, lexical and psycho-linguistic aspects and evaluated them using traditional machine learning models. The focus of this paper was to test domain-specific features on different types of classifiers first, and finally evaluating a pure multilingual approach on a combined English and Spanish dataset. The authors extended their experiment to a multilingual model design using less preprocessing and feature selection. Their study has demonstrated the importance of selecting domain-specific features in the domain of fake news identification. They concluded that it was possible to detect fake news spreaders based on a limited dataset of 300 Twitter users, by applying gradient boosting to a set of lexical, behavioural and psycho-linguistic features.

C. TARGET-BASED ACCOUNTS DETECTION

The target features identify end users that are affected by the fake news. A target can be a human, bot or cyborg depending on the nature and domain of fake news [70]. Although fake news can reach almost all the users through social media, an easy target will be those people that are more vulnerable and prone to get influenced by the fake news [88]. In order to understand and identify a potential victim of fake news, we can make use of victim dynamics feature as described below. Table 6 shows related studies which have made use of target victim features to detect fake news.

1) VICTIM DYNAMICS FEATURE

Victim dynamics mean thoroughly understanding the details of the end user. The details can include age, gender, education history, account creation history, network of followers, location etc. From the study [78], we find that generally new users with limited exposure to social media are targets of fake news spreaders, as they tend to believe anything presented to them due to lack of exposure. Teenagers and aged people

TABLE 5. Common propagator features.

Study	Propagation			Techniques		
	User Engagement	Time Dynamics	Platform-based Features	ML	DL	NLP
Kaliyar et al. (2021) [76]	✓	✗	✗	✗	✓	✗
Barhate et al. (2020) [95]	✗	✓	✓	✓	✗	✗
Sansonetti et al. (2020) [80]	✓	✗	✗	✓	✗	✗
De Nicola et al. (2021) [21]	✓	✗	✓	✓	✗	✗
Lu et al. (2020) [1]	✓	✗	✓	✓	✓	✗
Mendoza et al. (2020) [94]	✓	✗	✗	✓	✓	✗
Vogel et al. (2020) [69]	✓	✗	✓	✓	✗	✓
Pozzana et al. (2020) [84]	✓	✗	✗	✓	✗	✗
Bello et al. (2020) [70]	✓	✓	✗	✓	✗	✗

with limited knowledge of possibilities of fake news on social media are an easy target [73]. Similarly, people with low qualifications and coming from rural areas are more prone to be the victims of fake news [96]. Following papers have discussed the features of a target-based account. Yuan *et al.* [74] have proposed a human-based fake news detection technique SMAN which can detect fake news within 4 hours with an accuracy of over 91 percent, which is much faster than the state-of-the-art models. Chowdhury *et al.* [75] have proposed a credibility score-based model which detects the fake news by observing credibility of both publishers of the news and its users, making the model useful in fake news detection. Zhang and Ghorbani [30] propose a combination of cyborgs and bot detection scheme and discuss practical solutions versus research-based solutions. Ahmed *et al.* researched about who inadvertently shares fake news [97] and introduced a bias which enabled users to self report whether they shared deep fakes or not. The model also took into perspective the users who may have shared deep fakes and did not realize it. Albadi *et al.* [73] proposed a regression model which detect bots spreading hateful messages against various religious groups on Arabic Twitter. Ajesh *et al.* [88] detected fake news user profiles using random forest, optimised Naive bayes and support vector machine algorithms. Rodriguez-Ruiz *et al.* [67] took a hybrid, one-class classification approach to decide between bad bots and humans without the requirement of anomalous behavior examples. Shu and Liu [7] has addressed the increasing fake news propagation on social media and addressed the features of target victims that are more prone to be affected by fake news.

IV. DISCUSSION

Fake news dissemination and identification of fake news spreaders, propagators and targets is a challenging task. The use of Artificial Intelligence has proved fruitful in this regard. Most of the existing studies have used Deep Learning, Machine Learning and Natural Language Processing methods to detect fake news spreaders through feature extraction and classification.

Machine Learning is an application of AI that enables systems to learn and identify patterns which leads to decision-making without the intervention of humans. Machine Learning algorithms have specifically seen a boost in the field of fake news feature determination. During our extensive study, we have come across various ML algorithms that have been used for this purpose. The ML algorithms are trained using large datasets so that they can automatically detect the fake news spreaders [98]. Once fake news is shared on the internet, ML algorithms check its contents and detect fake news spreaders based on different features. Researchers have been trying to train machine learning classifier to detect with higher accuracy [99]. The better trained a classifier is, the more accurate it is [100]. Within ML framework, the common algorithms that have achieved better results include Neural Networks, Naive Bayes, Decision Trees and SVM.

Natural Language Processing deals with the interactions between computer systems and languages that enables computers to understand speech and text. NLP-based algorithms are used to detect linguistic and semantic patterns in fake news [101]. NLP supports AI in performing language related tasks such as creation of dialogues and interpreting words and sentences with ease [102]. Commonly used NLP techniques that achieved remarkable results compared to others include TF-IDF, LSA and LDA.

Deep learning is a branch of AI that comprises of artificial neural networks. The detection of fake news spreaders is complex and there are few shortcomings when it comes to using NLP alone for detection. DL and NLP techniques can be used in conjunction to improve automatic detection [103]. DL involves systematic representation of data and text analytics. The learning can be supervised or unsupervised. Some common DL techniques used for fake news spreaders detection with good performance include CNN, LSTM, BiGRU etc.

During our critical literature review, we found that the researchers have encountered many limitations out of which we have listed the most common challenges and future directions in our next section which can be addressed to build a more accurate fake news spreaders detection system.

V. RESEARCH CHALLENGES

Table 7 summarizes the most common challenges found by the researchers in creating an efficient fake news spreaders detection model.

TABLE 6. Common target features.

Study	Target	Techniques		
	Victim Dynamics	ML	DL	NLP
Yuan et al. (2020) [74]	✓	✓	✗	✗
Chowdhury et al. (2020) [75]	✓	✓	✗	✗
Zhang et al. (2020) [30]	✓	✓	✗	✗
Ahmedet al. (2021) [97]	✓	✓	✗	✗
Albadi et al. (2019) [73]	✓	✓	✗	✓
Ajesh Fet al. (2021) [88]	✓	✓	✓	✗
Rodr´ıguez-Ruiz et al. (2020) [67]	✓	✓	✓	✗
Shu et al. (2019) [7]	✓	✓	✗	✗

A. IDENTIFICATION AND DIFFERENTIATING A CYBORG WITH REAL USER

There is a kind of hybrid account known as cyborgs. The content is often new when a human takes over the bot account and the comments of that account at that point of time is authentic. They are better hider as robots and are pretty expensive to be made.[5] The cyborg activators mostly use the social media management platform Hootsuite, to simultaneously control multiple accounts at one time.[6] The challenge applies across all sorts of platforms, including and just not limited to Twitter, Facebook,[7] dating apps etc. Detecting a cyborg is not only difficult but also time consuming as they can hide behind the human's activities on the internet and have a very similar behavior to real time users. [104] Methods using other feature sets specially designed for them can contribute in detecting them.

B. LIMITATIONS IN TRAINING DATA

A large-scale dataset consisting of the real users and automated accounts including bot accounts is crucial in understanding relationships among different types of users, however, such datasets are limited and not updated. These datasets were built on relatively small data size, which can hardly generalize real-world scenarios, making it mostly unbalanced in real time. The effect of dataset size is more prominent in deep learning models. So, researchers should provide their datasets publicly so that other researchers can contribute to keep it updated [105]. Research that implements existing detection models and tests them on the some real public dataset is also needed. [23]

In addition, based on the studies we surveyed so far, we believe that the fake news detection lacks comprehensive dataset. Most of the datasets available, are based on political news. On the other hand, there are other aspects of social media concerning health, education, religion etc. Not much has been done in this regard. It is also challenging to create datasets of that caliber because of lack of accessibility to information of users due to privacy and confidentiality aspects. Finally, papers using hybrid-based fake news detection, such as [76] and [79] show that the accuracy of the experimental result when using both news content and social context, is more accurate as compared to using either news content or social context. However, collecting characteristics of the users can be very challenging and need regular updates to their status. Besides, Shu *et al.* [78] indicates that there are both explicit and implicit features contained in user profile and the implicit features such as personalities are very useful for analysis. The collection of implicit features of a large number of users would be even more difficult than the explicit features.

C. BIASES IN SURVEYS

One research challenge that was encountered is biases in the survey, and the way metrics were evaluated in the surveys created and conducted. For example, when researching who shared fake news, Apuke and Omar [91] only sampled people from Nigeria, which means that the data may not generalize to other countries. [97] introduced a bias in the way that users had to self report whether they shared deep fakes or not, meaning some users may have shared deep fakes and did not realize it.

D. MALICIOUS HUMAN AUTHORS

A fake news spreader can be a simple human user sometimes. This person can write their posts in such a way to avoid detection. They can meticulously craft a post which looks very real and not so different than a real post. In addition, a malicious actor may choose to wait some time before posting fake news. By doing this they may post regular content that goes undetected and in doing so gain real followers, friends, comments, and more. These metrics are all used by ML models to detect fake news accounts and posts. In the case of a human posting fake news and deliberately crafting posts maliciously, the challenge becomes much greater to detect it.

E. PLATFORM DIFFERENCE

Many papers focus on fake news detection method on a specific platform. For example, Twitter is chosen by many studies. The datasets used for their experiments are also Twitter users. As social media have various features and

[5]https://www.voanews.com/silicon-valley-technology/cyborgs-trolls-and-bots-guide-online-misinformation

[6]https://www.bbc.com/news/world-latin-america-42322064

[7]https://medium.com/@DFRLab/human-bot-or-cyborg-41273cdb1e17

TABLE 7. Open challenges in the detection of fake news spreaders.

Challenges	Causes
Cyborg Vs Real User	Cyborg are partially controlled by humans and partially by bots with fresh content on their profiles every now and then.
Limitations in training data	Most datasets are relatively small and are not comprehensive which impacts deep learning greatly. In addition, collecting user features, especially implicit features, is very challenging, and must be updated frequently.
Biases in Surveys	Surveys Conducted by other researchers can include biases when polling users
Malicious Human Authors	Malicious Humans are extremely hard to detect as they can create a legitimate account with good metrics and often wait to share fake news to fool ML models.
Platform difference	Most Research models focus on Twitter only for fake news detection meaning other platforms on which fake news is posted such as Instagram, Facebook may be hard to detect
Cross Platform	Fake news may be spread on lesser known social media platforms whereas legitimate users share on known ones

functions which are sometimes not similar to other social media;for example, Facebook and Instagram hosts a short story feature which exists for 24 hours and can be exploited to share fake news, whereas other platforms like Twitter just focus on feeds. So in order to detect fake news on other social media like Facebook, Instagram Youtube, Linked in, etc., the detection system may need some changes.

F. CROSS PLATFORM

Another major challenge in this area is number of available social media platforms besides the popular ones. Every region in the world has its own non popular social media platform which fake news spreaders target to spread fake news and then the genuine users take this information across different platforms, because they believe it to be true. It is highly unlikely to identify such accounts. There must be some cross platform control mechanism to verify links and articles before sharing them.

VI. FUTURE RESEARCH DIRECTIONS

From the survey, we conclude that malicious users and bot detection can be further improved by working on following areas.

A. PLATFORM INDEPENDENT CLASSIFIERS

Most of the literature focuses on detecting users and bots on Twitter platform, and most close-to-real datasets are normally available for Twitter platform. While Twitter is a popular platform, there are other platforms which are popular places of discussions for users around the world and so the spread of fake news is undiscovered and under researched on those platforms. A study [8] shows that Facebook has surpassed all other social media platforms for users now with WhatsApp and YouTube leading after it. Therefore, bots may have different features based on platforms and platform dependent models will render it impossible to detect bots on other platforms which are gaining popularity now. A potential direction would be to create platform independent datasets that can be used in building detection models that caters to other platforms.

[8]https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

B. MULTIPLE TYPES OF BOT DETECTION

The research shows that many bot detectors fail to detect other types of bot as bot masters are constantly changing features of bots making them more difficult to be detected online. A good future direction would be to have a classifier that could detect multiple type of bots separately instead of one. One way to do it can be to design an unsupervised method to cluster similar bot accounts based on dataset automatically, and then assign homogeneous accounts to the specialized bot classifiers. [66]

C. MULTILINGUAL DETECTION

In the literature, we can see a distinct lack of models trained in languages other than English. This presents a good future direction of study. If a model was trained with numerous languages, then it could be generalized to other countries that have a different native language. The style of fake news and how it is written may differ country to country as well, so a dataset from a country speaking that language would be a good contribution, instead of translating existing datasets into other languages. Some research have detected multi lingual satire detection [106], others have detected general fake news in English, Spanish and Portuguese only. [107]

D. REAL TIME DETECTION

The research also shows models which detect fake news bots after they have posted fake news on the internet. A future direction which researchers may choose, is a real time detection. This means that a social media platform can implement a real time model, which would flag users and posts when they try to post fake news. Even with existing system such as the one Varshney and Vishwakarma [108] describes, it is only described as a real time system with limited abilities to detect fake news spreaders in real time.

E. COLLECTION OF IMPLICIT FEATURES OF USERS

Implicit features are not directly shown from user profiles. Shu *et al.* [78] shows that implicit features perform better than

explicit features. However, implicit features are hard to obtain because they are the inference of user behaviors. For example, the inference of the personality of a user is hard, specially when that user does not have too much information on his account. If people start inferring the implicit features, this may result in biasness. One future research direction would be finding a good way of extracting both explicit and implicit features from user accounts.

F. SOCIAL BOT WITH GOOD USES

Bots are used as a way of spreading fake news. They can heavily influence on human behavior by manipulating their emotions. For example, bots can create fake news with content which look true to invoke human fear or surprise of a fake fact or a group of people. Then, they would be more likely to share this news with others. This is how attackers are using bots to spread fake news. One future direction is that whether it is possible for bots to be used to encourage people or spread facts and positive emotions to make the society better. One direction of future research could be figuring out how can bots be used for positive utilities and conducting a study on cyborg coordination and communication.

VII. CONCLUSION

OSNs has nowadays become the most integral part of everyone's life, and has become the major source of information for everyone. While it has lot of benefits, it has also shown some serious drawbacks in the form of spread of fake news, done to manipulate users minds and their decisions [41]. Both human and bots share fake news, and the bots can mimic human features very closely. There are numerous challenging issues which currently require further investigation such as differentiating a user account from automated accounts. In this survey, we have reviewed all the state of art methods used by researchers in detecting a malicious human user, and a bot based on source-based, propagator-based and target-based features of user accounts. Lastly, we have mentioned common challenges and future research directions from our survey which will help future researchers come up with a sophisticated classifier with better accuracy rate.

REFERENCES

[1] Y.-J. Lu and C.-T. Li, "GCAN: Graph-aware co-attention networks for explainable fake news detection on social media," 2020, *arXiv:2004.11648*.

[2] C. Grimme, M. Preuss, L. Adam, and H. Trautmann, "Social bots: Human-like by means of human control?" *Big Data*, vol. 5, no. 4, pp. 279–293, Dec. 2017.

[3] C. Snider. (Sep. 2021). *Social Media Statistics*. [Online]. Available: https://chrissniderdesign.com/blog/resources/social-media-statistics/

[4] Y. Roth and D. Harvey, "How Twitter is fighting spam and malicious automation," *Twitter [Blog], June*, Jun. 2018.

[5] K. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, and F. Menczer, "Arming the public with artificial intelligence to counter social bots," *Hum. Behav. Emerg. Technol.*, vol. 1, no. 1, pp. 48–61, Jan. 2019.

[6] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 36–211, 2017.

[7] K. Shu and H. Liu, "Detecting fake news on social media," *Synth. Lectures Data Mining Knowl. Discovery*, vol. 11, no. 3, pp. 1–129, 2019.

[8] R. R. Mourão and C. T. Robertson, "Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information," *Journalism Stud.*, vol. 20, no. 14, pp. 2077–2095, Oct. 2019.

[9] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2931–2937.

[10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.

[11] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," 2017, *arXiv:1707.07592*.

[12] C. Wardle and H. Derakhshan, "Thinking about 'information disorder': Formats of misinformation, disinformation, and mal-information," in *Journalism,'Fake News' & Disinformation*, C. Ireton and J. Posetti. Paris, France: Unesco, 2018, pp. 43–54.

[13] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," *Proc. Comput. Sci.*, vol. 141, pp. 215–222, Jan. 2018.

[14] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 11, no. 1, 2017, pp. 280–289.

[15] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proc. 26th Int. Conf. World Wide Web Companion (WWW) Companion*, 2017, pp. 963–972.

[16] M. Spradling, M. Allison, T. Tsogbadrakh, and J. Strong, "Toward limiting social BotNet effectiveness while detection is performed: A probabilistic approach," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2019, pp. 1388–1391.

[17] S. Cresci, M. Petrocchi, A. Spognardi, and S. Tognazzi, "Better safe than sorry: An adversarial approach to improve social bot detection," in *Proc. 10th ACM Conf. Web Sci.*, Jun. 2019, pp. 47–56.

[18] K. Warwick, "Cyborg morals, cyborg values, cyborg ethics," *Ethics Inf. Technol.*, vol. 5, no. 3, pp. 131–137, 2003.

[19] S. Castillo, H. Allende-Cid, W. Palma, R. Alfaro, H. S. Ramos, C. Gonzalez, C. Elortegui, and P. Santander, "Detection of bots and cyborgs in Twitter: A study on the Chilean presidential election in 2017," in *Proc. Int. Conf. Hum.-Comput. Interact.*, Cham, Switzerland: Springer, 2019, pp. 311–323.

[20] T. Khaund, B. Kirdemir, N. Agarwal, H. Liu, and F. Morstatter, "Social bots and their coordination during online campaigns: A survey," *IEEE Trans. Computat. Social Syst.*, early access, Aug. 19, 2021, doi: 10.1109/TCSS.2021.3103515.

[21] R. D. Nicola, M. Petrocchi, and M. Pratelli, "On the efficacy of old features for the detection of new bots," *Inf. Process. Manage.*, vol. 58, no. 6, Nov. 2021, Art. no. 102685.

[22] A. Balestrucci and R. D. Nicola, "Credulous users and fake news: A real case study on the propagation in Twitter," in *Proc. IEEE Conf. Evolving Adapt. Intell. Syst. (EAIS)*, May 2020, pp. 1–8.

[23] M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of bots in social media: A systematic review," *Inf. Process. Manage.*, vol. 57, no. 4, Jul. 2020, Art. no. 102250.

[24] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, May 2021, pp. 255–261.

[25] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[26] C.-C. Wang, "Fake news and related concepts: Definitions and recent research development," *Contemp. Manage. Res.*, vol. 16, no. 3, pp. 145–174, Sep. 2020.

[27] J. M. Burkhardt, "History of fake news," *Library Technol. Rep.*, vol. 53, no. 8, pp. 5–9, 2017.

[28] S. B. Naeem, R. Bhatti, and A. Khan, "An exploration of how fake news is taking over social media and putting public health at risk," *Health Inf. Libraries J.*, vol. 38, no. 2, pp. 143–149, Jun. 2021.

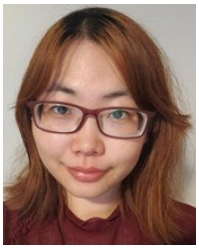
[29] J. Prier, "Commanding the trend: Social media as information warfare," *Strategic Stud. Quart.*, vol. 11, no. 4, pp. 50–85, 2017.

[30] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102025.
[31] S. Hakak, W. Z. Khan, S. Bhattacharya, G. T. Reddy, and K.-K. R. Choo, "Propagation of fake news on social media: Challenges and opportunities," in *Proc. CSoNet*, 2020, pp. 345–353.
[32] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The limitations of stylometry for detecting machine-generated fake news," 2019, *arXiv:1908.09805*.
[33] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96–104, Jul. 2016, doi: 10.1145/2818717.
[34] R. Gorwa and D. Guilbeault, "Unpacking the social media bot: A typology to guide research and policy," *Policy Internet*, vol. 12, no. 2, pp. 225–248, Jun. 2020.
[35] R. Var ao, "Patrícia Campos mello (2020), the hate machine: Notes from a reporter on fake news and digital violence [a máquina do ódio: Notas de uma repórter sobre fake news e violência digital]. são paulo: Companhia das letras. language: Portuguese Brazilian. ISBN-10: 853593362x. PBK, 296 pages," in *Digital War*, vol. 2. Brazil: Springer, 2021, pp. 96–97.
[36] N. M. Lee, "Fake news, phishing, and fraud: A call for research on digital media literacy education beyond the classroom," *Commun. Educ.*, vol. 67, no. 4, pp. 460–466, Oct. 2018.
[37] F. W. Judith, S. C. Baraka, G. Gregory, and K. Joseph, "Clickbait-style headlines and journalism credibility in sub-saharan Africa: Exploring audience perceptions," *J. Media Commun. Stud.*, vol. 13, no. 2, pp. 50–56, May 2021.
[38] A. Devitt and K. Ahmad, "Sentiment polarity identification in financial news: A cohesion-based approach," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 984–991.
[39] A. Olteanu, E. Kıcıman, and C. Castillo, "A critical review of online social data: Biases, methodological pitfalls, and ethical boundaries," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Feb. 2018, pp. 785–786.
[40] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 675–684.
[41] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106983.
[42] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020.
[43] G. Baym and J. P. Jones, *News Parody Political Satire Across Globe*. Evanston, IL, USA: Routledge, 2013.
[44] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A multi-modal framework for fake news detection," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2019, pp. 39–47.
[45] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.
[46] R. Toews. (May 2020). *Deepfakes are Going to Wreak Havoc on Society. We are not Prepared*. [Online]. Available: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to%-wreak-havoc-on-society-we-are-not-prepared/?sh=731ebdc47494
[47] A. Tsakalidis, "Misinformation in Eu elections 2019: A post analysis," Int. Hellenic Univ., Thessaloniki, Greece, Tech. Rep. 11544/29544, 2020. [Online]. Available: https://repository.ihu.edu.gr/xmlui/handle/11544/29544
[48] D. Freelon and C. Wells, "Disinformation as political communication," *Political Commun.*, vol. 37, no. 2, pp. 145–156, 2020, doi: 10.1080/10584609.2020.1723755.
[49] K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu, "User identity linkage across online social networks: A review," *ACM SIGKDD Explor. Newslett.*, vol. 18, no. 2, pp. 5–17, 2017.
[50] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," 2019, *arXiv:1905.12616*.
[51] Á. Figueira and L. Oliveira, "The current state of fake news: Challenges and opportunities," *Proc. Comput. Sci.*, vol. 121, pp. 817–825, Jan. 2017.
[52] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning," in *Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT)*, Sep. 2019, pp. 1–4.
[53] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, "A new approach to bot detection: Striking the balance between precision and recall," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 533–540.
[54] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins, "#FluxFlow: Visual analysis of anomalous information spreading on social media," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1773–1782, Dec. 2014.
[55] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. (IMC)*, 2007, pp. 29–42.
[56] S. Gianvecchio, Z. Wu, M. Xie, and H. Wang, "Battle of botcraft: Fighting bots in online games with human observational proofs," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, 2009, pp. 256–268.
[57] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?" in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 620–627.
[58] B. 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," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 11, no. 1, 2017, pp. 759–766.
[59] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in *Proc. 5th Int. Workshop Adversarial Inf. Retr. Web*, 2009, pp. 41–48.
[60] X. Zhou and R. Zafarani, "Network-based fake news detection: A pattern-driven approach," *ACM SIGKDD Explor. Newslett.*, vol. 21, no. 2, pp. 48–60, 2019.
[61] S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, pp. 1–23, Jul. 2017.
[62] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, pp. 1146–1151, May 2018.
[63] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on Twitter," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 153–164.
[64] C. Braker, S. Shiaeles, G. Bendiab, N. Savage, and K. Limniotis, "Botspot: Deep learning classification of bot accounts within Twitter," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Cham, Switzerland: Springer, 2020, pp. 165–175.
[65] J. N. Paredes, G. I. Simari, M. V. Martinez, and M. A. Falappa, "Detecting malicious behavior in social platforms via hybrid knowledge- and data-driven systems," *Future Gener. Comput. Syst.*, vol. 125, pp. 232–246, Dec. 2021.
[66] M. Sayyadiharikandeh, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "Detection of novel social bots by ensembles of specialized classifiers," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2725–2732.
[67] J. Rodríguez-Ruiz, J. I. Mata-Sánchez, R. Monroy, O. Loyola-González, and A. López-Cuevas, "A one-class classification approach for bot detection on Twitter," *Comput. Secur.*, vol. 91, Apr. 2020, Art. no. 101715.
[68] M. Cardaioli, M. Conti, A. D. Sorbo, E. Fabrizio, S. Laudanna, and C. A. Visaggio, "It's a matter of style: Detecting social bots through writing style consistency," in *Proc. Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2021, pp. 1–9.
[69] I. Vogel and M. Meghana, "Detecting fake news spreaders on Twitter from a multilingual perspective," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2020, pp. 599–606.
[70] H. R. M. Bello, L. Heilmann, and E. Ronan, "Detecting fake news spreaders with behavioural, lexical and psycholinguistic features," in *Proc. CLEF (Working Notes)*, 2020, pp. 1–12.
[71] M. Lichouri, M. Abbas, and B. Benaziz, "Profiling fake news spreaders on Twitter based on tfidf features and morphological process notebook for pan at clef 2020," Thessaloniki, Greece, Tech. Rep., 2020. [Online]. Available: http://ceur-ws.org/Vol-2696
[72] G. Lingam, R. R. Rout, D. Somayajulu, and S. K. Das, "Social BotNet community detection: A novel approach based on behavioral similarity in Twitter network using deep learning," in *Proc. 15th ACM Asia Conf. Comput. Commun. Secur.*, Oct. 2020, pp. 708–718.
[73] N. Albadi, M. Kurdi, and S. Mishra, "Hateful people or hateful bots?: Detection and characterization of bots spreading religious hatred in Arabic social media," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, pp. 1–25, Nov. 2019.
[74] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu, "Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning," 2020, *arXiv:2012.04233*.
[75] R. Chowdhury, S. Srinivasan, and L. Getoor, "Joint estimation of user and publisher credibility for fake news detection," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 1993–1996.
[76] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, Feb. 2021.

[77] R. K. Kaliyar, A. Goswami, and P. Narang, "EchoFakeD: Improving fake news detection in social media with an efficient deep neural network," *Neural Comput. Appl.*, vol. 33, pp. 8597–8613, Jan. 2021.
[78] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2019, pp. 436–439.
[79] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: A tool for fake news collection, detection, and visualization," *Comput. Math. Org. Theory*, vol. 25, no. 1, pp. 60–71, Mar. 2019.
[80] G. Sansonetti, F. Gasparetti, G. D'aniello, and A. Micarelli, "Unreliable users detection in social media: Deep learning techniques for automatic detection," *IEEE Access*, vol. 8, pp. 213154–213167, 2020.
[81] H. Telang, S. More, Y. Modi, and L. Kurup, "Anempirical analysis of classification models for detection of fake news articles," in *Proc. IEEE Int. Conf. Electr., Comput. Commun. Technol. (ICECCT)*, Feb. 2019, pp. 1–7.
[82] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 312–320.
[83] Y. Long, *Fake News Detection Through Multi-Perspective Speaker Profiles*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017.
[84] I. Pozzana and E. Ferrara, "Measuring bot and human behavioral dynamics," *Frontiers Phys.*, vol. 8, p. 125, Apr. 2020.
[85] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment aware fake news detection on online social networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2507–2511.
[86] J. C. S. Reis, A. Correia, and F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar./Apr. 2019.
[87] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 797–806.
[88] F. Ajesh, S. Aswathy, F. M. Philip, and V. Jeyakrishnan, "A hybrid method for fake profile detection in social networkusing artificial intelligence," in *Security Issues and Privacy Concerns in Industry 4.0 Applications*. Hoboken, NJ, USA: Wiley, 2021, pp. 89–112.
[89] Y. Wu, Y. Fang, S. Shang, J. Jin, L. Wei, and H. Wang, "A novel framework for detecting social bots with deep neural networks and active learning," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106525.
[90] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007.
[91] O. D. Apuke and B. Omar, "Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users," *Telematics Informat.*, vol. 56, Jan. 2021, Art. no. 101475.
[92] S. Martinez-Bea, S. Castillo-Perez, and J. Garcia-Alfaro, "Real-time malicious fast-flux detection using DNS and bot related features," in *Proc. 11th Annu. Conf. Privacy, Secur. Trust*, Jul. 2013, pp. 369–372.
[93] R. U. Rahman and D. S. Tomar, "New biostatistics features for detecting web bot activity on web applications," *Comput. Secur.*, vol. 97, Oct. 2020, Art. no. 102001.
[94] M. Mendoza, M. Tesconi, and S. Cresci, "Bots in social and interaction networks: Detection and impact estimation," *ACM Trans. Inf. Syst.*, vol. 39, no. 1, pp. 1–32, Jan. 2021.
[95] S. Barhate, R. Mangla, D. Panjwani, S. Gatkal, and F. Kazi, "Twitter bot detection and their influence in hashtag manipulation," in *Proc. IEEE 17th India Council Int. Conf. (INDICON)*, Dec. 2020, pp. 1–7.
[96] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, "State of the art models for fake news detection tasks," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 519–524.
[97] S. Ahmed, "Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size," *Telematics Informat.*, vol. 57, Mar. 2021, Art. no. 101508.
[98] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, and M. S. Choi, "Detecting fake news using machine learning : A systematic literature review," 2021, *arXiv:2102.04458*.
[99] E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in *Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC)*, Jun. 2019, pp. 1–5.
[100] P. Ksieniewicz, M. Choraś, R. Kozik, and M. Woźniak, "Machine learning methods for fake news classification," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, Cham, Switzerland: Springer, 2019, pp. 332–339.
[101] C. Busioc, S. Ruseti, and M. Dascalu, "A literature review of NLP approaches to fake news detection and their applicability to romanian-language news analysis," *Transilvania*, vol. 10, pp. 65–71, Oct. 2020.
[102] A. Hamid, N. Shiekh, N. Said, K. Ahmad, A. Gul, L. Hassan, and A. Al-Fuqaha, "Fake news detection in social media using graph neural networks and NLP techniques: A COVID-19 use-case," 2020, *arXiv:2012.07517*.
[103] D. S and B. Chitturi, "Deep neural approach to fake-news identification," *Proc. Comput. Sci.*, vol. 167, pp. 2236–2243, Jan. 2020.
[104] S. Qi, L. AlKulaib, and D. A. Broniatowski, "Detecting and characterizing bot-like behavior on Twitter," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling Predict. Behav. Represent. Modeling Simulation*. Cham, Switzerland: Springer, 2018, pp. 228–232.
[105] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on Twitter," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 185–192.
[106] G. Guibon, L. Ermakova, H. Seffih, A. Firsov, and G. L. Noé-Bienvenu, "Multilingual fake news detection with satire," in *Proc. CICLing, Int. Conf. Comput. Linguistics Intell. Text Process.*, 2019, pp. 1–12.
[107] H. Q. Abonizio, J. I. de Morais, G. M. Tavares, and S. Barbon, Jr., "Language-independent fake news detection: English, portuguese, and Spanish mutual features," *Future Internet*, vol. 12, no. 5, p. 87, May 2020.
[108] D. Varshney and D. K. Vishwakarma, "Hoax news-inspector: A real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles," *J. Ambient Intell. Humanized Comput.*, vol. 12, pp. 8961–8974, Nov. 2020.

WAJIHA SHAHID was born in Karachi, Pakistan. She received the master's degree in computer science information technology from NED University Pakistan, with a focus on machine learning and NLP. She is currently pursuing the Ph.D. degree in computer science from the University of New Brunswick, with a focus on fake news and fake news spreaders detection using machine learning and NLP.

She has five years of work experience as a Software Quality Assurance Engineer.



YIRAN LI is currently pursuing the B.Sc. degree in computer science with UNB, with a focus on cybersecurity. She is currently doing co-op work as a Full Stack Developer.

DAKOTA STAPLES is currently pursuing the B.C.S. degree in cyber security and the Honors Designation. He is pursuing to continue his studies and obtain his M.C.S. degree as well. His research interests include phishing/spam detection and NLP.

GULSHAN AMIN received the bachelor's and master's degrees in computer networks in India. She is currently pursuing the M.S.C. degree in cyber security with the University of New Brunswick, Canada. She worked as a Lecturer with the College of Computer Science and Information Technology, Shaqra University, Saudi Arabia. She has vast teaching experience due to having worked in various educational institutions locally and abroad.

SAQIB HAKAK is currently an Assistant Professor with the Canadian Institute for Cybersecurity (CIC), Faculty of Computer Science, University of New Brunswick (UNB). He is having more than five years of industrial and academic experience. His current research interests include risk management, fake news detection using AI, security and privacy concerns in IoE, applications of federated learning in IoT, and blockchain technology. He has received the number of gold/silver awards in international innovation competitions and is serving as a technical committee member/reviewer of several reputed conference/journal venues.

ALI GHORBANI (Senior Member, IEEE) has held various academic positions for the past 40 years and is currently a Professor of computer science, the Tier 1 Canada Research Chair in cybersecurity, and the Director of the Canadian Institute for Cybersecurity, where he established, in 2016. He was the Dean of the Faculty of Computer Science, University of New Brunswick, from 2008 to 2017. He has spent over 25 years of his 40-year academic career carrying out fundamental and applied research in machine learning, cybersecurity, and critical infrastructure protection. He is the co-inventor on three awarded and one filed patent in cybersecurity and web intelligence and has published over 300 peer-reviewed articles during his career. He has supervised over 200 research associates, postdoctoral fellows, and students during his career. His book, *Intrusion Detection and Prevention Systems: Concepts and Techniques* (Springer, October 2010). He developed several technologies adopted by high-tech companies and co-founded three startups, Sentrant Security, EyesOver Technologies, and Cydarien Security, in 2013, 2015, and 2019, respectively. His current research interests include network information security, machine learning, complex adaptive systems, and critical infrastructure protection.

Dr. Ghorbani served as the Co-Editor-In-Chief for the *International Journal of Computational Intelligence Systems*, from 2007 to 2017.