Fake news Impacts and approaches to detect fake news

Anwar Hashem[1], Maryam Yousif[1], Nouf Saif[1], Safa Faisal[1], Sanaa Elyassami[1]
[1]Department of Information Security Engineering Technology
Abu Dhabi Polytechnic, Abu Dhabi, UAE

Abstract. Fake news troubles the current world due to its disruption of attitudes, beliefs and values for ulterior purposes. This phenomenon is enabled by the continued surge in digital communications that have improved the speed of communication as well as accessibility ton news sources. This paper reviews scholarly literature on fake news by looking at fake news impacts, detecting approaches and shows the difference between machine learning and artificial intelligence. Understanding the fake news phenomenon is crucial to establishing effective means of detecting and weeding out fake news and reiterating the credibility of news on both mainstream and digital media platforms.

Keywords: Fake news, Detection, Mainstream media, Machine learning, Social media, Journalism, Deep learning, Naïve Bayes, Artificial Intelligence, Computational linguistics.

1. Introduction

Interests in studying phenomenology around the creation and dissemination of fake news have risen significantly over the past decade. These interests coincide with the introduction and continued proliferation of social media platforms and internet communications across the world. Allen et al. [1] highlight the number of scholarly publications that contain the words "fake news" on Google Scholar. 2210 publications with the title "fake news" have appeared since January 2017 while only 73 publications have the same in the years leading 2017. This statistic shows the growing sensitivity of Fake News not only as an academic topic but also as a social concern worth studying. Moreover, the concentration of fake news on social media platforms as opposed to traditional mainstream platforms. According to the Google Scholar compilation, 308 articles had the title "television" news while 708 had the terms "online news", "social media" or just news [1]. As such, the increased popularity of fake news corresponds to the increased dependency on social media platforms for news. According to some study, 62 percent of adults in the US depend on social media for their news with Facebook and Twitter being the most popular news sources.

This literature review highlights the problems associated with the dissemination of fake news and how different platforms

can leverage technology to limit the spread of fake news.

2. Research Method

2.1 Research Objective

The study's objective is analysis fake news impacts, detecting approaches and comparing the machine learning with artificial intelligence. This was realized by performing a systematic literature review to determine previous research on evaluating news reliability. During the systematic literature review, 13 relevant studies were used to show the impact of fake news on the field of journalism and impacts on the society. Further, the study discusses different approaches to fake news and compare machine learning with artificial intelligence

2.2 Search Terms

There were multiple search terms used to facilitate the findings of pertinent journal articles like:

("Detect" OR "Find" OR "Identify" OR "Understand")

AND

("Fake news" OR "Mislead online news" OR "False news" OR "Misinformation")

AND

("Social media" OR "Social Networking sites" OR "Social networkers" OR "Journalists")

AND

("Machine learning" OR "Deep learning" OR "Artificial Intelligence")

2.3 Selection Criteria

Inclusion Criteria. The selected studies met the listed criteria:

1. The studies had been published amid 2015-2020.

2. The articles have been published in English with a core focus on fake news detection and its impact.

3. The selected articles were to be found in reputable journals related to technology and articles cited multiple times.

4. The articles had vital terms such as "fake news," online news", "social media," or "machine learning," amongst other keywords identified in the search terms section.

Exclusion Criteria. The studies that were excluded followed the listed criteria:

1. The articles that had not been published online such as books and newspapers.

2. Fake news articles but not related to information technology.

3. Records that have been published on untrusted websites.

2.4 Flowchart of Search Process
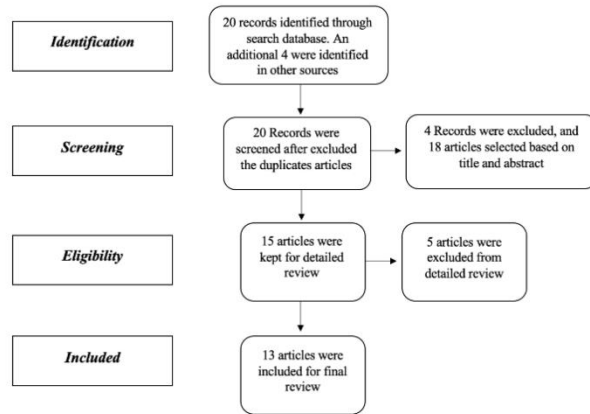
3. Findings

3.1 Impacts of Fake News on Journalism as a field of practice

The perspectives from different scholars prove that fake news remains an inherent problem in both mainstream and social media. Their impacts on journalism as a career, on the society, and news consumption cultures draw critical insights that need to be understood before identifying solutions to the problems. Express the concern of fake news dissemination the credibility of journalism in mainstream media. In a study, reveal that 0. 1 percent of Twitter users generate 80 percent of fake news on the platforms [2]. The volume of fake news only serves to overshadow the credibility of mainstream news outlets and therefore generates flawed perceptions of news outlets. On credibility, Vergeer [3] conducted a study to determine how frequent journalists verify online news

before accepting or proceeding to develop articles on them. According to Vergeer's findings, only 36.8 percent of journalists in the Netherlands verify online news sources while 0.5 percent never attempt to verify online news. These statistics could be negative or positive in different countries. Nonetheless, they all point to the fact that fake online news has a negative impact on the careers of journalists and the reputation of mainstream media houses.

			Percentage	Cumulative Percentage
0	Lowest	Never in all of the three presented cases	0.5	0.5
1			0.6	1.1
2			0.9	2.0
3			3.3	5.3
4			4.1	9.3
5			7.4	16.7
6			13.2	29.9
7			16.1	45.9
8			17.3	63.2
9	Highest	Always in all of the three presented cases	36.8	100.0
			100.0	100.0

N = 666, median = 8

Figure 1: Levels of verification of online information [3].

3.2 Impact of Fake News on Society

Fake news has an adverse impact on society and could be used to misguide the fights against vices like human trafficking "the fight for women's equality, the environmental crisis, and a myriad of political, social, economic, and cultural issues" [4]. About two out of three (64%) adults in USA say fabricated news stories cause a great deal of confusion about the basic facts of current issues and events. Although Americans see fake news as

causing a great deal of confusion in general, most are at least somewhat confident in their own ability to identify when a news story is almost completely made up. About four-in-ten (39%) are very confident, while other 45% are somewhat confident. Only 9% are not very confident, and 6% are not at all confident [4]. This is similar to American's general faith in their ability to tell when online information is trustworthy. However, fake news challenges require specific approaches that delve into the heart of the problem in both traditional and digital media platforms. Developing a model to detect fake news can help in addressing these effects. To start with, by minimizing fake news, the model will enhance the authenticity balance of the news ecosystem. In addition, the model will reduce the possibility of people being manipulated to accept false or biased beliefs. Currently, propagandists have created bots for spreading fake news. Besides, the model will improve the way people interpret and respond to news. Essentially, people are more likely to make the right decisions if they are supplied with authentic information. Furthermore, the model will improve governance and national security.
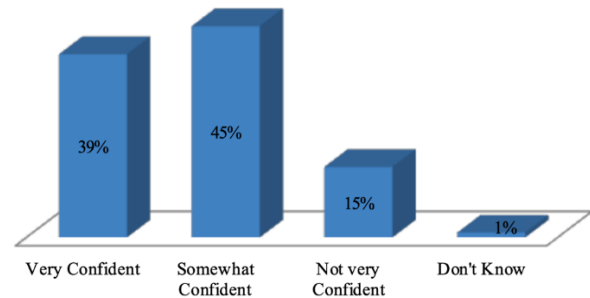

Figure 2: Many American believe that fake news is sowing confusion [4].

3.3 Fake News Detection on Social Media using Deep Learning

There has been a lot of development of the communication through network. And some people use this communication as a way to change the information. Therefore, fake news has become a trend to these people. The purpose of the fake news it that it potentially causes problems to the government or between people which then leads them to make some serious problems to the society. Since the internet is easy to access and it has wide variety of information sources it has become the most important thing to the people. People usually think that everything in the website is being true, however in 2016 of the election campaigns they have noticed a lot of fake news spreading and people believing in it. Some of the big companies (social media) have noticed the danger of these fake news so they began to work on something that will detect the fake news on their platforms [5].

However, this technique of detecting the fake news is hard and still very complicated. Here where we will begin to understand more about the deep learning.

BERT is a language model made by scientists at Google which depends on transformers. The system to follow with BERT falls under transforms learning. BERT is given as of now pretrained on a huge book corpus (books, Wikipedia, and so forth.) with the point that the last client plays out a fine-tunning stage to adjust the model to his particular issue. It input data format is different from the ones used for the other two architectures since it is based only on text strings. The word tokenization and separation processes are already included in the input data function for this model. The experience picked up during the improvement of this models permits us to express that the utilization of profound learning models for this errand can be possibly advantageous for a wide scope of entertainers, from informal organization organizations to the last client so as to relieve the expanding duplicities on the Internet.
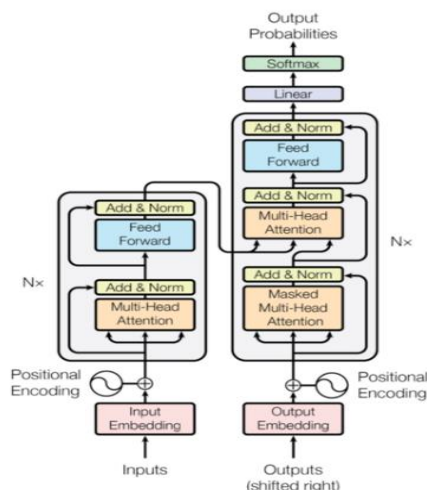

Figure 3: *Transformer block diagram* [5].

3.4 Fake News Detection on Social Media using Geometric Deep Learning

Social media is a double-edged sword for news consumption, but they are one of the main news sources, on account of offering low cost, easy access, and rapid dissemination. Spreading news through social media also comes with the risk of exposure to 'fake news' that contains poorly checked or intentionally false information to mislead and manipulate the readers for an economic or political benefit. Detecting news automatically poses challenges such as news interpretation requires the knowledge of political or social context or 'common sense', which current natural language processing algorithms are still missing.

The geometric deep learning approach for fake news detection was trained and tested on news stories, verified by professional fact-checking organizations, that were spread on Twitter. the experiments indicate that social network structure and propagation are important features allowing highly accurate (92.7% ROC AUC) fake news detection [6]. It is a learning fake news specific propagation patterns by exploiting geometric deep learning, a novel class of deep learning methods designed to work on graph-structured data. Geometric deep learning naturally deals with heterogeneous data, thus carrying the potential of being a unifying framework for content, social context, and propagation-based approaches.

3.5 DEAN: Learning Dual Emotion for Fake News Detection on Social Media

Microblogging is a way for people to post, share and seek information because of its easy to use, however it leads us to easily create fake news which has some societal consequences. That's why detecting fake news on microblogs is highly important. Some publishers intend to publish fake news in order to make the users angry or make them do things they shouldn't do. This is where it effects our society. In this article I have learned how to dual emotion for the fake news detection.

As you can see in the figure 4, it shows how the people reacts to the news either they're emotionally sad or mad, and because of these small facts it can lead to many problems. Emotions is more viral not only in the real world but also on social media.
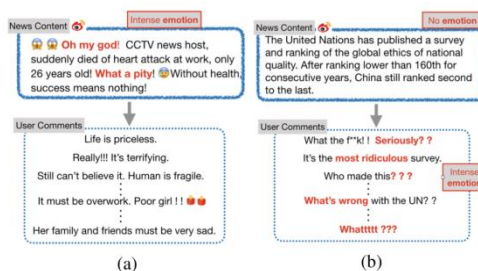

Figure 4: Fake news posts [8].

The publishers usually post news with intense emotions which can trigger a high level of physiological arousal in the crowd. Or may present the news objectively to make it convincing but with controversial content which evokes intense emotion in the public, encouraging it spread- ing widely. For us to maintain this problem Our solution to these two challenges results in a novel Dual Emotion-based fAke News detection framework (DEAN). This strategy could be effectively applied via web-based media by utilizing on the web feeling corpus, which could generally beat the constraints of sen- timent word references.

3.6 Comparison of Various Machine Learning Models for Accurate Detection of Fake News

Fake news consists of news that is not well researched or deliberate steps have been taken to spread misinformation or hoaxes via different forms of news distribution networks. Using a computational model of probabilistic and geometric machine learning models has used to tackle this issue. Comparing different vectorizers using various classifiers to predict the fake news helped to find the appropriate vectorizer for fake news detection.

Among all the combinations that have been tested, SVMs along with TF-IDF Vectorizer gave the highest accuracy and produce better results than other models when trained on larger current datasets. The logistic regression is another model worth noting. It performs equally well for both Count and TF-IDF Vectorizer and hence may be a better choice than SVMs. The logistic regression and SVM models result in better scores with larger datasets, which make it difficult to predict which one will execute better in the future. Where Naive Bayes and Decision Tree did not improve scores by a considerable margin with an increase in sample size. However, Naive Bayes gave the highest accuracy scores for smaller datasets

and hence has considerable importance. Decision Trees on the other hand performed poorly throughout and hence seem to be a bad candidate for fake news classification. Better results were expected from neural networks, they performed the poorest of all and consistently results in very low accuracy. Using other text classification methods such as n-grams and linguistic analysis may give better results with neural networks.

Fake news classification is majorly limited by the lack of detailed and proper datasets for testing models. Thus, models that can make better predictions with larger datasets are the ones that need focus. The same models can be trained using larger sample sizes and accuracy scores need to be judged again. The classification was focused on the news content along with the title, source of news, and engagement statistics for the highly accurate classifications.

	Naive Bayes	SVM	Log. Reg.	Dec. Tree	NN
Count Vectorizer	0.863	0.891	0.916	0.825	0.499
TF-IDF Vectorizer	0.854	0.928	0.910	0.816	0.499

Table 1: *Comparison of the accuracy of fake news classification using different classifiers* [9].

3.7 Computational Linguistics

Blackwell [10] contrived an innovative insight that will effectively contend fake news. They have exhibited a combination of computational linguistics and fact-checking in 2017. This approach offers automated fake news detection performance that discriminates between real and fake news. to be quantitatively characterized. Moreover, provenance will be approached with reference to the historical evaluation of the author and the proliferation of computational linguistics and pattern spread. Computational linguistics includes judgment on lexical, syntactic, and semantic levels of items in a piece of news. Moreover, it functions in a way that testers have observed that the improved system was able to detect fake news with an accuracy of up to 76%, which is close to human performance. Linguistic features and fact-verifying go hand in hand on detecting fake news because not only lexical analysis should be considered, but also the authenticity of a piece of news, meta-features, and model features progressed by machine learning are necessary. Computational linguistics entails the researchers to present RPI software intending to develop the software rating to locate event stories as GTD is the combination. The author has examined a scoring scale for detecting deceptive

information depending on the markers, as demonstrated in figure (5). If a story or a piece of news is too adorned or embellished, the deviation value will grow. Hence, the markers will increase, and vice versa. Furthermore, a story or a piece of news is also considered as a deviation measure if it repeats a certain point within the content three times. Contrarily, if a news article neglects some keys of information within the contents, the markers will decrease.
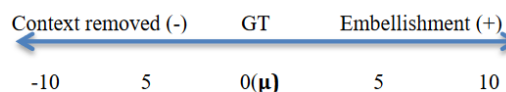

Figure 5: Scoring Scale [11].

This scoring scale demonstrates how directions indicate the deviations from the original event. For instance, if a news article is specific or appointed, the deviation score will be -10. However, excessive and embellished news articles will score the deviation score as +10. The way that the scoring scale evaluates the result is that it detects these news articles in terms of an average of parts of speech, tones, ratios, etc. Therefore, the overall score becomes a label used for both storage and signal identification. Storage labels are useful for comparisons used in source provenance and for news articles that generate without the GT sources.

3.8 Deep Hybrid Model

However, Ruchansky [11] believes that constructing some features of fake news using a social graph is time-consuming, as it demands a mind that profoundly understands artificial knowledge. Additionally, she assumed that some detection methods like knowledge-based and content-based cannot integrate the characteristics of text, response, and source accurately. For this reason, those authors have concocted the idea of a deep hybrid model for detecting fake news elastically, which comprises two modules. In formula (11), it illustrates the entire architecture process. The first module is the capture module. It is used to extort all the feedback text information of content and low-dimensional descriptions of articles and users. Then, they get assembled by LSTM.

$$\mathbf{x}_t = (\eta, \Delta t, \mathbf{x}_u, \mathbf{x}_\tau)$$

Formula 1: Capture Model Formula [11].

where η assigns to the number of subscriptions, x_u assigns to the user's global characteristics, and x_τ represents the text characteristics of all reply comments.
The second module is the Score module. It forms a user relationship network and diminishes the latitude to calculate a pair of s_i and $\hat{y}_i$. The authors have decided to use the matrix built by the number of users to subscribe to the number of topics that participate in the calculation. Then, it performs SVD reduction to gain the assignment of y_i of the user and cooperate in the calculation, and then pass them to an m_j mask processing. According to the author, this architecture detects deceptive information through using the description of the article reached from the content of the users' participation, as well as the description of the article taken from the network constructed by the users' behavior.

3.9 Machine Learning VS Artificial intelligences battle

Fake news has influenced the internet globally to the extent that it is unattainable to verify the truthfulness of a piece of news, or whether the user who has published is a human or a bot. This phenomenon has become inventive that a fake news article contains overexpressed, exaggerated, and manipulated content. Authors argue that there is more than one approach to depend on that will utterly solve this phenomenon. Some of them assume that applying artificial intelligence may structure a robust fake news detection, and some of them opposed to this hypothesis and sanctioned that machine learning is the way to go. Thus, we

have proposed a comparison between both approaches, concluding with our observations.

3.9.1 Artificial Intelligence (AI)

The first group of authors asserted that artificial intelligence (AI) is capable of detecting the deception of information through pattern recognition efficiently, as it is known for its ability to separate the good from the bad. Those authors believe that embedding artificial intelligence tools into fake news detection will render the system more powerful than it is. That is because AI differentiates between fake news and real news by taking a cue from articles that people see fallacious in the light of checking disinformation or misinformation when it is deployed speedily.

That group of authors also insisted on how artificial intelligence will make a major difference to social media, and that it is the savior of the phenomenon of fake news spread globally. They have addressed that AI can quickly find words or patterns of words that are likely to interpret fake news stories. Moreover, it can discover the meaning of an entire web article utilizing natural language processing (NLP). In other words, it goes deeply though the subject of a news article published on verified websites.

Then, it checks its headline, main body text, and the geo-location and ascertains if other websites conveyed the same information.

Nevertheless, we have explored that malicious users and bots go for visual media to express fake news by manipulating a sophisticated image, video editing applications, and automated manipulation algorithms that provides complex editing that is too difficult to detect by an abstract eye or even with current image analysis and visual media forensic tools. There should be more effort to be exerted in order to have a system that detects the veracity of news comprehensively.

Authors and Mark Zuckerberg (a fake news detector) have predicted that AI can potentially meet that comprehensive fake news detection [12]. They also have added to their prediction that AI will be the solution of spreading deceptive information digitally by granting programs that battle the computational propaganda. Also, it provides powerful tools split between content-based and response-based identification that combat suspicious content. It works like a human fact-checker by matching articles with verified and trusted references to spot disinformation.

3.9.2 Machine Learning

The versus group of authors believe that detecting news using the machine learning approach will be more practical and feasible in this case, as many programmers have already constructed such detection. Those authors proposed two features that will make fake news detection more accurate, news gathering and complex news handling. Machine learning is more of a predictive approach, in which it collects news or information that hold class attributes with its respective values depending on the news itself whether it is true positive, true negative or false positive, false negative. This will improve confusion production through attributes like precision. This operation will be handled by news gathering feature. Once this operation is successfully accomplished, the model containing gathered news will go onto the next feature.

In the complex news handling feature, the system follows multiple steps in order to analyze complex news that is uninterpretable to traditional models: Tokenizing, Padding, Encoding, embedding matrix formation, Model Formation, Model Training, and predicting the model. Tokenizing begins first with parsing text to remove words in the input news presented in the LIAR dataset. Then, these words will be encoded and lemmatized to integers or floating-point values so that they are inserted into a machine learning algorithm. In fact, the dataset includes 10,000 different news articles along with their class attributes [13]. Next up is padding. This stage uses pad_sequences() function to pad variable-length sequences with 0.0 as a default value. As soon as the value is specified through the "value" argument, the value may change.
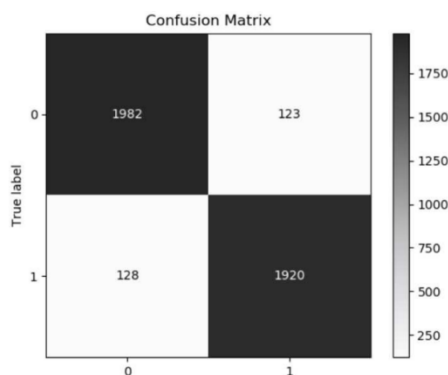

Figure 6: LTSM [13].

3   Conclusion

This review of literature samples scholarly opinions on different aspects of fake news including impacts on journalism and society and approaches to detecting fake news that can be used to suppress the dissemination of fake news. However, the phenomenology of fake news frequently changes as different social media platforms come into being, and the number of users on existing sites

continues to bulge. As such, the burden of fake news is greater than ever. There is a need to clearly define the boundaries of fake news building upon the abstract definitions of the term, as highlighted in this literature review. To pursue the problem of fake news, one needs to establish contexts and the succeeding mechanisms for detecting and eliminating fake news in these contexts.

References:

[1] Allen, Jennifer, et al. "Evaluating the fake news problem at the scale of the information
  ecosystem." *Science Advances* vol. 6, no.14, 2020, eaay3539.

[2] Tsfati, Yariv et al. "Causes and Consequences of Mainstream Media Dissemination of Fake
News: Literature Review and Synthesis." *Annals of the International Communication Association* 44.2 (2020): 157–173.

[3] Vergeer, Maurice. "Incorrect, Fake, and False. Journalists' Perceived Online Source Credibility and Verification
Behavior." *Observatorio (OBS*)* 12.1 (2018): n. pag. Web.

[4] Nagi, Kuldeep. "New social media and the impact of fake news on society." *ICSSM Proceedings, July* (2018): 77–96.

[5] Rodríguez, Álvaro Ibrain, and Lara Lloret Iglesias. *Fake News Detection Using Deep
Learning*. 2019.

[6] Monti, Federico et al. *Fake News Detection on Social Media Using Geometric Deep
Learning*. 2019.

[7] Monti, Federico. "Geometric Deep Learning: Approaches and Applications." 22 May 2015.

[8] Guo, Chuan et al. *DEAN: Learning Dual Emotion for FAke News Detection on Social Media.*
2019.

[9] Poddar, Karishnu, Geraldine Bessie Amali D, and Umadevi Ks. "Comparison of Various
Machine Learning Models for Accurate Detection of Fake News." *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*. IEEE, 2019.

[10] Blackwell, Adam. "Op Ed-Opinions And Editorials-Educators, Not Engineers, Should Lead
 The Fight Against Fake News". *Against The Grain*, vol 29, no. 6, 2017. *Purdue University (Bepress)*, doi:10.7771/2380-176x.7880.

[11] Ruchansky, Natali et al. "A Hybrid Deep Model For Fake News Detection". *CSI*, vol 4, no.
 4, 2017, Accessed 27 Sept 2020.

[12] Abiodun, Alao. "How Artificial Intelligence Tools Are Deployed In The Fight Against Fake
 News". *The Nation*, vol 4, no. 4, 2020.

[13] Sharma, Nikhil. "Fake News Detection Using Machine Learning". *International Journal Of
  Trend In Scientific Research And Development (IJTSRD)*, vol 4, no. 4, 2020.