

# The Misinformation Era: Review on Deep Learning Approach to Fake News Detection

K.B.R. Perera  
IT17134286  
Department of Software Engineering  
Sri Lanka Institute of Information  
Technology  
Malabe  
it17134286@my.sliit.lk

**Abstract**—Public refer to the media as the fourth power after the executive, legislative and the judicial. It is of utmost importance to a nation that the public should get the true information from their media whether it is traditional or nontraditional media. “Fake News” the term was popularized during the 2016 US presidential election period refers to the deliberately misinforming the public against or favor of a person or entity. As those news are a threat to the national discourse and national security research on the subject matter has increased. We looked at those researches that have been done to identify fake news manually and automatically using deep learning techniques. Through that we have identified some research gaps that we presented in this paper. As the “Fake News” term is too vague and has multiple definitions we go through those to identify different levels of fake news. We also looked at social media handling of the fake news pandemic and identified the new pathways for future research.

**Keywords**—*Fake news, Deep learning, Detection Algorithms, Social Media*

## I. INTRODUCTION

“Fake News” is a term used most often these days by politicians and media personnel. Fake news most often described as fabricated or made-up story regarding any political or non-political issue with intent to mislead or deceive. These types of stories can be distributed by the traditional news media like print papers, television as well as non-traditional media such as social media and websites. These news or stories mainly focused on the political landscape to tarnish a reputation of an entity or a person, change the public discourse to a desirable place for these news distributors. And this is a great threat to the free discourse of ideas, debate and the democracy itself. Also some of these fake news intend to gain popularity and sensationalism within the masses.

For example in the 2016 US presidential elections there was a widely believed right wing conspiracy named “pizzagate”. During the presidential election Wikileaks, which is an organization dedicated to bringing out secrets of governments all around the world, dumped millions of emails that have been accessed through Hillary Clinton’s campaign manager John Podesta’s email. This email dump created a lot of news cycles as it is dumped by batches every single day until the election day. People who proponents of this conspiracy theory believed that in the emails there are coded messages that are connected to sex trafficking ring with the democratic party officials. They believed one of the alleged establishments of this ring was a pizzeria in

Washington DC in the US. This conspiracy theory was popularized by the right wing media outlets and personal, with the social media platforms like 4chan, 8chan, facebook and twitter. As a result one man from the state of North Carolina traveled to the pizzeria in washington with a rifle to investigate and fired shots inside the pizzeria. Also the manager and staff were threatened by the conspiracy theorists. This conspiracy has been debunked by many independent sources including the Washington DC police.

Another incidence in US at 2019, Covington Catholic High School students with navite american man caught attention of the media. First in a short clip showing the one of the student and navite american man in standoff that was uploaded to twitter. Mainstream media has picked up on that video and made it a newscycle to bash the students as they wore hats regarding one of the right wing political movements in the US. The school also apologized to the natvive american man as the video shown to be students is disrespected towards him. This also proved the apparent political bias of the US mainstream media as they did not fact check the story before running it. After the media’s irresponsible handling of the incident, students were getting death threats and the school was closed due to the fears about student safety. After this situation and a few days later another video has emerged that shows the full story of what happened in that situation which shows that the native american man was the one that started the confrontation with the students and not the other way around. Mainstream media was keen on the misrepresentation of the incident at the first place as that would be in line with their political ideals. After the full video has been released the Covington school students have filed a defamation lawsuit against the few mainstream media companies ultimately resulting in undisclosed settlement with CNN and The Washington Post.

Deep learning was popularized around 2012 with a highly influenced paper of imagenet convolutional neural networks for object recognition[1]. Deep learning is a concept based on the brain neurons and work similar to them. Deep learning neural networks use several hidden layers inside the network for training. That would increase the learning parameters which will provide a better model for the problems we tend to tackle. As deep learning is used for pattern recognition, based on sample data it can be used to recognize the unverified facts and options of news articles. Using different datasets researchers have been proposing a lot of models, algorithms to recognition of these articles and statements. We go through the research done in the topic, and identify the research gaps. Identifying fake

news is not easy as it has many different definitions. We tried to identify different types of fake news and identify whether it should be banned on public forums. As most of the fake news is on social media, we discussed the fake news in the context of social media and problems facing.

As shown by the examples this pandemic of fake news stories come from across all political factions and social factions. So that the public should be able to access the news without any editorialisation and any misinformation. And the news websites in social media should be regulated in a fair manner. These regularizations should happen in a transparent manner so that the users do know the social media platform itself is not biased toward an opinion. Twitter has faced allegations being biased towards a one political faction which would not happen if the algorithms that used for fake news were public and transparent. Also these detection algorithms are essential to the free discourse of ideas and debate. Otherwise the public would not be able to filter through the fake news to get to the real news which will ultimately decide the fate of a nation. These detection algorithms can be used as a weapon against free speech if used in an unfair manner.

## II. WHAT IS FAKE NEWS?

There are few types of fake news or misinformation based stories. Let explore those fake news types[2]. Some inteded as humor and fool people and some are deliberately produced for harm to a person or an entity. It is important to recognize the difference between comical or unintended mistakes against harmful fake news. This should be considered when building a model to identify fake news because otherwise any artistic works could be identified as fake news.

### A. Satire or Parody

These types of stories are not deliberately created to cause harm. These are made to fool the person reading or sharing. A parody is a comedic commentary that imitates a personality or a work. Parodies of public figures exist on twitter and other social media. Satire on the other hand uses public figures and creative works for delivering a message to the audience. This message can be politically motivated and satire is not a commentary on the work but the scapegoat to delivering a message to a certain group or the public at large.

### B. False Connection

When in a news stories media such as images and audio, and captions does not match the story this makes a false connection between the real story and the sources presented. Also mainly headlines could be altered in a way to attract more readers by saying outrageous claims. This can also be used as a clickbait technique for news sites as the headlines are the first thing any reader would read and see. These headlines of the stories can have potential connection to actual news stories presented but can be used to get views to the actual story using the media worthiness of the altered headlines.

### C. Misleading content

Misleading content is when an information is used as a manner of which to frame an individual or an issue that can not be provided as evidence. These types of news can make false claims based on the real information to deliberately

misinform the reader. If the overall impression left by a news story is based on the unsubstantiated and unproven claims, that can be categorized as misleading content.

### D. False Context

Information about the news story is correct but the information and the sources is represented in a way that is not accurate. Context of the story rather matters than the information about the story. This also could happen due to the political or other biases of the writer. Maybe even if the information was right, it does not have any relevance to the actual story and those information was presented in such a way to harm or promote an entity or a person.

### E. Imposter content

News stories under this category try to impersonate some genuine news organization, news personal or entity. This can be used to gain trust of the reader by pretending to be a trustworthy publisher or author. Even fake news websites can impersonate other genuine news websites. These types of stories contain false facts which they claim to come from genuine sources which also are made up or been impersonated.

### F. Manipulated content

If an information or media about a news story is manipulated in such a way to harm or promote a person or an entity, that type of stories can be categorized as a manipulated content. Sometimes videos can be edited by a speaker, to show what they meant was different. Also this can be done in quoting a person which puts a part of the actual statement rather than the full statement. That way the writer or the publisher can be content false with intent to deceive.

### G. Fabricated content

All of the content of this type of news stories are fabricated and trying to purposefully deceive and harm. Content of the stories contains made up evidence, facts and sources that persuade the reader to think it is an actual news story. Anyone can write whatever that comes to their mind on social media and share, might potentially be shared by other users without any verification. Also some news sites leaning in some political direction like left-wing or right-wing could create fake news stories to get clickbaits and to tarnish their opposition. These stories are really dangerous to the public trust on the institutions and functionality of democracy itself.

Even though researchers categorized fake news based on the severity of some content being fake, there are other definitions that could interpret fake news differently some times in a biased manner. There is no clear cut definition of fake news and that may be why the term is misused. Also the term itself is too vague that could cause confusion among the general public. Maybe a general standard should be introduced to identify clearly what fake news is and what is not. This could be some guidelines that the reader can be followed to identify whether the story is genuine.

### III. FAKE NEWS DETECTION IN SOCIAL MEDIA AND PROBLEMS

In the social media context only there are some different algorithms at play. Fact Checking algorithms, Rumor Detection algorithms, Stance Detection algorithms and Sentiment Analysis algorithms are few of those. Fact checking algorithms check truthfulness of the speech or any other means of communication on the platform. These algorithms mainly focus on political speech of politicians and activists. Rumor detection is done to identify any statement that consists of unverified information that can not be verified in any way. Stance detection is used to identify whether a user is on what side of a debate from the text of the statement. Tracking the emotion of a user about any particular topic and whether they are impressed or not impressed by the statement is sentiment analysis. Those algorithms and others check the users behaviors, feelings and combine those to determine if the statement is fake or not [3].

Almost all the social media networks consist of Application Programming Interface(API) for developers so that they can build interactive apps that get the features of the social media. That means programmers can build bots to act like a real person in the platform. Since the bot automatically does what it has been programmed to do, malicious programmers can add fake news to the platform without even without being on the platform itself. For example twitter bots can look at a phrase in a tweet and retweet it, comment on it and many more. This issue was also raised in the 2016 US presidential election as intelligence agencies concluded that twitter and other social media bots and fake profiles did influence the election.

As for why fake news in social media is more problematic than the news websites, as most people around the world do not deal with news websites rather go through their feed to gain information. When some users are particularly targeted with fake news it could impact the lives of the victims in a much more vicious way than a random website publishing fake information. In seconds or minutes some fake articles, fake statements could go viral and get millions of shares. Even if the corrections are made after the fact that would not get as much attention as the article or statement with the misinformation. It is evident that today's social networks all around the world consist of these things. This is why the fake news detection algorithms and using those to detect fake news early as possible are important.

Trusting algorithms over people is somewhat problematic in the mind of the general public as they can make mistakes on large scales. This is why some of the social media companies implemented an appeal process for the detection of any policy violations. Machines do get it wrong sometimes and the appeal process could correct it. Yet the algorithms that run those calculations do not have any oversight to see whether those algorithms could be biased towards opinion vs facts. If these algorithms were actually in the public domain, the public could verify, and even improve those algorithms for better. And again even the appeal process can be biased as most people on social media platforms do not believe the social media itself has the best interest for the users. That may be why conspiracy theories against social media companies start and gain popularity.

### IV. LITERATURE REVIEW

From 2015 onwards there is a lot of research interest in the topic of "Fake News" as the term was popularized through the US presidential election campaigns in 2016. Papers are mainly focused on the history of the subject matter and how it will affect the lives of the people today going forward. Also there are a lot of discussions on how to identify those news stories with the help of computers. Lot of papers mainly focused on social media as the platform to identify the fake news, as they are the ones people get their dose of daily news from.

One of the most cited papers on the topic [4] discussed why this is an important topic on the internet age. Paper also shows that fact checking on social media would sometimes be counter productive against the journey for truth with political or social bias at play. This paper also states how to combat against the fake news on the internet, by computer algorithms and other methods of actions such as using fact checking websites, with a call to action for the more research of the topic. Paper[4] on fundamentals of fake news states that the difference between malicious users that create the fake news and normal users which are just sharing and posting is blurred due to the knowledge level of the normal users. Detection strategies for the fake news article have been discussed here which includes style-based fake news detection, propagation-based fake news detection and credibility-based fake news detection. Deep learning and fact checking websites have been suggested as the potential resources for the detection of the fraudulent news stories. In a study[6] conducted on college students in the US, to identify the fake vs real news stories it was found that the student did correctly identify whether the weather news was fake or real for 62% of the time. This study used 63 subjects with more than 700 observations that concluded that nearly 40% of news stories were wrongly identified by the students. However this study also concluded that the students are going to use more verification strategies such as using google to search about the relevant topic and search other sources to verify the story, than the general public comparing other research to this study.

Detection of fake news, researched for obtaining a model which will successfully could be implemented in social media and search engines alike. Models have been proposed using artificial intelligence, speaker profiles, sentimental analysis, natural language processing and deep learning with many technologies. Paper[7] on the model of fake news detection using speaker profiles argue that the normal methods using only text and media does not become successful because the author of the news stories and how he or she views the world does not count into the detection model. By attaching the party affiliations, credit history, location of the residence and many other user specific variables, they were able to achieve almost 15% increase to the best models using the orthodox methods. This model archives the hybrid model by attaching said speaker's information to the basic classification model. But the shared news on social media and many of the news sites on the internet do not have well recognized reporters which would become impossible to gain prediction using this model alone. With privacy concerns that people have it makes it hard to acquire the necessary data for this model to be implemented in large scale. If we can implement a model for identifying an author or speaker, overtime we could build up a model generated profile that could eventually be fed into

the proposed model in this paper. Another paper[8] which proposed a model for social media fake news detection that builds on an algorithm using what is called a “tri-relationship” between the publisher, news stories and the users of the platform. They proposed to have a social score like system for the social media users which calculate by their previous behaviors and whether they are intentionally distributing the fake news stories. When a story is shared, the model does evaluate the publisher as well as the content of the news, then the person who shared the news evaluates by the model, which determines whether the news story is fake or not. Users' personal scores will be considered for classification model. However this also has the previously stated problems with publishers / authors also the social media score system will not be sensitive to parody like commentary which does not have intention to mislead.

Datasets are very important to detect these types of problems. They need to be thoroughly classified and should easily be accessed by the public. Liar dataset is such a great dataset which would help the research on the topic by a substantial amount. Author of the paper[9] collected more than 12000 stories over the decades that have manually labeled whether a news story is false or not. Also paper proposed a model with the dataset, hybrid convolutional neural network to integrate metadata with text. That would increase the accuracy of text only deep learning models with paper benchmarks for the dataset. In a paper[10] published in 2017, proposed detection models for social media which combat against the fake news through multiple models. These models are Knowledge-based model which is a model based on known facts, Style-based model which is a model based on writing style of the news story, Stance-based model which is a model based on determining the writer is favor or against some target entity and lastly Propagation-based model which is a model on the relevance of the same topic posts and whether the a post is related and accurate to the original story. Interestingly the paper also discussed the Echo chamber effect on social media and how it affects the individuals to determine whether a story is false or not.

3HAN another deep learning model[11] proposed to identify fake news using only text based inputs. This mode has three levels of architecture, for each word, sentence and headline with a news article vector to represent an article in a hierarchical bottom-up manner. When an article feeds into the model it gives the probability of it being fake as an output then would use human or automatic fact checkers to verify the result obtained by the model. Model trained on a dataset of fake articles, from 2016-2017 with 65% of them being US presidential news, 15% for both regional news and world news and 5% of entertainment news. Using text only analysis they achieved 96% of accuracy of fake news detection. But this would mean this model is not capable of capturing manipulated media content and non-text information presented on an article. Unique approach to the problem in hand given by this paper[12] by a simple model which will output whether a claim is genuine or not. Feed the claim and the body of text to the model which then identifies features of an article and tries to find related articles. Then the model will evaluate and determine if the related article is agreeing or disagreeing with the claim that was put into the model in the first place. But the accuracy of the most is yet to be improved and model only process text based input only.

Deep learning based model proposed in another paper[13] called CSI model. They recognized the main three characteristics to evaluate text of an article, users promoting the said article and what type of response the article is getting from the other users in the platform. Model they proposed consists of modules Capture, Score and Integrate where the name comes from. Text of the news story gets captured and gets a score based on the responses to that article, and these two are integrated to determine if the story is actually false or not. But since this is based on user reports and comments, this could potentially be used as a censoring tool by the malicious users or an entity.

## V. LIMITATIONS OF FAKE NEWS DETECTION

Even though we have technologies to identify fake news at a certain pace it is not sufficient in any manner because it becomes more challenging as technologies advance and evolve over time. Thus fake news circulation becomes much more accelerating. This makes the detection algorithms obsolete if they cannot detect the fake content before it becomes a viral thing. False and unproven news does circulate at a higher rate than a normal news story which is also a challenging factor when it comes to tracking and detecting fake news.

Deep learning algorithms too have limitations, even newly proposed algorithms such as Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN) often struggle with identifying fake news content because those fake stories are written in a way to resemble a true news story. Also the fact checking needs to be done as the text analysis alone could not prove a news story is fake or not definitely. Methods that use feedback from other users in social networks do not have the capabilities to identify fake news stories without reasonable feedback from the platform users. Linguistic analysis-based methods also suffer from the amount of information they are able to capture, and can be used to determine the outcome.

In a recent paper[14] that studied the limitation of fake news detection algorithms found that the neural language models (LMs) have the ability to bypass the detection algorithms and generate news that is unable to be identified as fake news with the current technologies. Currently, neural language models are being improved rapidly, and fake news concerns emerged as deep learning fake news detection algorithms could not keep up with it. In this study they test the models with the human written text and machine generated text. Even though models did fairly well with the human written text for fake news detection, machine generated text was not able to be detected by deep learning algorithms.

## VI. RESEARCH GAP

Currently there is a lot of research going on in the field of deep learning generally. Most of those advancement archived by researchers directly going to contribute to the algorithms that used to detect the fake news, since accuracy of those algorithms and models most depend on deep learning research. But there are some areas that a research gap exists on this specific topic of fake news detection algorithms.

There is no standard for news to identify as fake news or not. But there are some proposed categorizations of fake news. If there is a standard available to any reader of a news story, they can use those standards to evaluate it.

Nevertheless most people do not check the validity of the information presented or and find other resources to verify the first story. This type of standard can be helpful to the detection models and would improve the research so that standard should work as the benchmark for the models. This also could help create unbiased social media platforms that are not skewed towards a political party, positions or and opinions. That could also be one part of the solution to the polarization of politics all around the world. It would also help to regularize social media giants like facebook, twitter to have free speech on their platform without making biased decisions when it comes to the fake news. Even though we have to fight against fake news on the internet and elsewhere it should not infringe on freedom of speech, free discourse as that is the basis for modern democracy and nations.

Protection of sources is pretty much essential for the news stories that are regarding national security or any other important matter that is contentious. These news stories can be labeled as fake news by the government or any other entity which has been accused or proved of any wrongdoing. These detection algorithms also should look into protecting the publisher, author and sources of an actual story. This cannot be done using profiling the publisher, authors and his or her sources. That would raise privacy concerns about the data of authors which will ultimately potentially expose the real sources of a news story. This is much concerned with the freedom of press, as some governments tried to spy on journalists to gather information. Many of the western governments also do this practice as much as developing world countries. Researchers should be more concerned about the data protection as much as the gathering of data which is not researched thoroughly when it comes to fake news detection models.

As previously mentioned there is a lack of good dataset to run the experimental research on. So if research wants to progress at a brisk pace, the public dataset needs to be increased. Some researchers have introduced large datasets for fake news detection algorithms though the number of datasets must be increased and the size of data should be larger.

Another research issue is that there is no wiggle room for jokes, parodies and satire. Researchers did not give it much consideration as their main goal is to identify the false information. Those parody and jokes should be considered with their artistic values. Some websites are dedicated to this kind of work and such a website is babylonbee.com which is making fun of politicians, public figures and many others. Those types of work should not be identified as deceiving the public and should be treated like an art form. If a fake news detection model cannot identify the difference between parodies/satire and fake news, it could lead to massive confusion regarding the algorithms. Human intervention is needed to determine whether the news story is a parody or satire which is a path for more research. Models should be intelligent enough to identify these subtle differences and more research is needed to achieve that.

## VII. CONCLUSION

Deep learning algorithms is the future of fake news detection. But the detection algorithms must be improved to detect edge case scenarios such as parody and satire. There are several research gaps including the lack of datasets, protection of sources in news stories and world standard of fake news detection. Future work and research must

consider those factors of research gaps with improving the current understanding of fake news. Social media giants like facebook, twitter should expose their detection algorithms to the public domain so that the users of the platform trust those algorithms to make correct decisions. Otherwise users will assume there is bias against certain groups and political opinions which will ultimately create more fake news. We should always fight against fake news as that is detrimental to society.

## REFERENCES

- [1] G. Marcus, "Deep Learning: A Critical Appraisal," arXiv.org, 02-Jan-2018. [Online]. Available: <https://arxiv.org/abs/1801.00631>. [Accessed: 11-Aug-2020].
- [2] O. B. e. Caucaso, "Fake news. It's complicated," Media Freedom Resource Centre OBCT. [Online]. Available: <https://www.rcmediafreedom.eu/Publications/Reports/Fake-news.-It-s-complicated>.
- [3] R. Oshikawa, J. Qian, and W. Y. Wang, "A Survey on Natural Language Processing for Fake News Detection," arXiv.org, 05-Mar-2020. [Online]. Available: <https://arxiv.org/abs/1811.00770>. [Accessed: 10-Aug-2020].
- [4] D. M. J. Lazer et al., "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, Mar. 2018, doi: 10.1126/science.aao2998.
- [5] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "Fake News," presented at the WSDM '19: The Twelfth ACM International Conference on Web Search and Data Mining, Jan. 2019, doi: 10.1145/3289600.3291382.
- [6] C. Leeder, "How college students evaluate and share 'fake news' stories," *Library & Information Science Research*, vol. 41, no. 3, p. 100967, Jul. 2019, doi: 10.1016/j.lisr.2019.100967.
- [7] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake News Detection Through Multi-Perspective Speaker Profiles," *ACL Anthology*. Available: <https://www.aclweb.org/anthology/I17-2043>.
- [8] K. Shu, S. Wang, and H. Liu, "Beyond News Contents," presented at the WSDM '19: The Twelfth ACM International Conference on Web Search and Data Mining, Jan. 2019, doi: 10.1145/3289600.3290994.
- [9] W. Y. Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," arXiv.org, 01-May-2017. [Online]. Available: <https://arxiv.org/abs/1705.00648>.
- [10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Sep. 2017, doi: 10.1145/3137597.3137600.
- [11] Singhanian, S., Fernandez, N., & Rao, S. (2017). 3HAN: A Deep Neural Network for Fake News Detection. In *Neural Information Processing* (pp. 572–581). Springer International Publishing. [https://doi.org/10.1007/978-3-319-70096-0\\_59](https://doi.org/10.1007/978-3-319-70096-0_59)
- [12] Singhanian, S., Fernandez, N., & Rao, S. (2017). 3HAN: A Deep Neural Network for Fake News Detection. In *Neural Information Processing* (pp. 572–581). Springer International Publishing. [https://doi.org/10.1007/978-3-319-70096-0\\_59](https://doi.org/10.1007/978-3-319-70096-0_59)
- [13] N. Ruchansky, S. Seo, and Y. Liu, "CSI," presented at the CIKM '17: ACM Conference on Information and Knowledge Management, Nov. 2017, doi: 10.1145/3132847.3132877.
- [14] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The Limitations of Stylometry for Detecting Machine-Generated Fake News," *Computational Linguistics*, vol. 46, no. 2, pp. 499–510, Jun. 2020, doi: 10.1162/coli\_a\_00380.