

FAKE MEDIA DETECTION BASED ON NATURAL LANGUAGE PROCESSING AND BLOCK CHAIN APPROACHES

**Mrs.C.Archana¹ Assistant Professor¹, P.Shalini², P. Akshaya Nandini³, K.Madhuri⁴, K.
Revanth⁵**

archanachinna5491@gmail.com ¹, shalinipappula5555@gmail.com ², nandiniakshaya68@gmail.com ³,
Kyamamadhuri666@gmail.com ⁴, kusumarevanth8@gmail.com ⁵

Department of CSE, Teegala Krishna Reddy engineering college, Telangana, India

Abstract: Social media network is one of the important parts of human life based on the recent technologies and developments in terms of computer science area. This environment has become a famous platform for sharing information and news on any topics and daily reports, which is the main era for collecting data and data transmission. There are various advantages of this environment, but in another point of view there are lots of fake news and information that mislead the reader and user for the information needed. Lack of trust-able information and real news of social media information is one of the huge problems of this system. To overcome this problem, we have proposed an integrated system for various aspects of blockchain and natural language processing (NLP) to apply machine learning techniques to detect fake news and better predict fake user accounts and posts. The Reinforcement Learning technique is applied for this process. To improve this platform in terms of security, the decentralized blockchain framework applied, which provides the outline of digital contents authority proof. More specifically, the concept of this system is developing

a secure platform to predict and identify fake news in social media networks.

Index Terms : Natural language processing , blockchain , fake media , reinforcement learning

1. INTRODUCTION

Variety of shared information is the realistic part of social media. From 2017, fake news has become a very considerable topic until now, which 365% frequently used online [1]. Struggling with fake news becomes an unsolved problem in social networks in the data and information consumption application layer and becomes a serious and challenging issue in information advancement that appears in diplomatic, economic, and political sectors. The fake information revelation point to the unnecessary process in the network resources. Moreover, it contains the content totality and validity based on the available service [2]. Therefore, the wrong information sharing relevance the Quality of Trust (QoT) to apply on the news distribution [3]. Machine learning text classification improves the level of security that is needed in social media daily-based networking.

Expressing feelings or sharing an opinion through the social networking portal from the non-government organization's survey contains many fake accounts and information circulating the portal based on a suitable channel. In this case, the harmful and unwanted accounts need to pass from the network to give more space to the data center and manage the mess and political problems in the network.

Another related area for information extraction is propaganda which is special for political purposes [4]–[6]. The fake news forging language is very crafty in terms of predesignate to arouse and aggravate the emotion of users for spreading fake information [7]–[9]. Fake news detection is the capability of contents analysis based on truth in the shared information [10]–[12]. Along with the number of noisy and unstructured data, growth of the number of users, and news, there is a need for an automatic solution for extraction of fake news [13]–[15]. These terms become limited based on the recent developments in machine learning, deep learning, and artificial intelligence. Proofing the digital contents authorship is one of the mandatory steps for information sharing. To do this, blockchain is a suitable and promising framework that is the decentralized and secure platform to improve fake information extraction. A blockchain system continuously increases the number of blocks which each block has the previous block cryptographic hash, timestamp and transactions information [16]. The data integrity is guaranty with the blockchain and all the transnational information store in it. This aspect of blockchain makes it a famous platform in this approach. As a case study, we collected the social media contents from Facebook and Twitter, which are famous information sharing platforms with thousands of users that upload millions of daily news

and posts on various topics. This research aims to authorize fake users and information using the blockchain, NLP, and machine learning techniques. More specifically, the proposed system is the preventative approach based on the integrated techniques for the concept of fake data extraction combining with gamification components. Reinforcement learning is the learning-based algorithm that improves the system quality based on the provided information. If the information is wrong, the system prevents using similar information as before to reduce the fake and wrong information rating.

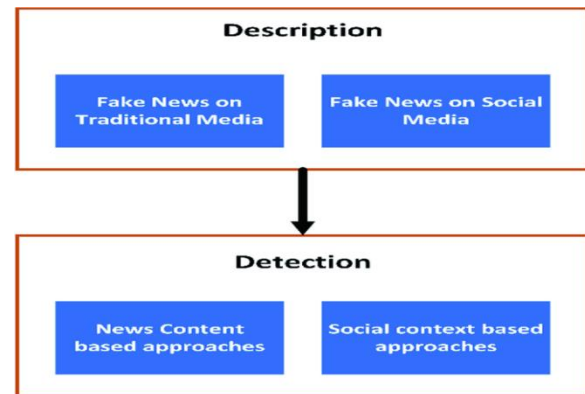


Fig 1 Example Figure

The main contribution of this paper is threefold: Designing the fake news prevention system instead of a detection system and applying the Natural Language Processing (NLP) for the detailed text analysis based on the shared contents. Applying the proof of authority protocol and designing financial roots. This process is the strong aspect of this system to find fake user information and accounts. Applying the Reinforcement Learning technique for predicting the learning rate of the system and extracting fake accounts. Finding the relationship between contents, extracting the similar meaning and structure of the shared information to avoid sharing fake news. It

shows the overview of the proposed approach in fake news detection based on the integrated method. The applied blockchain system is permissioned network that every participants are supposed to register and required authentication to make them qualified to join to blockchain network. In permissioned blockchain only authenticate users have allowance of joining to network which this process is the responsibility of user identification manager. This process also required the authentication certificate and enrollment for the valid participants. The aim of the proposed system is to store the news data in the distributed ledger which is reliable and secure platform.

2. LITERATURE SURVEY

Transformations of professional political communications in the digital society (by the example of the fake news communication strategy):

The report is focused on the analysis of revolutionary changes taking place in the field of professional political communications of the digital society. The purpose of this research is to study the specific features of the emergence and evolution context of the “Fake news” communication technology. The methodological basis for the analysis is the multiparadigm approach, which allows to combine heuristic potentials of the system-sociological methodology, interactive approach and theory of communicative action. Secondary sources devoted to the problems of modern political communications, as well as thematic content of social networks such as Twitter, Facebook, VKontakte, etc., were the empirical base. The following results were obtained during the performed analysis. Firstly, in recent years the toolkit of modern political communications has

been expanding, which testifies to the development of a well-known “electronic or monitoring democracy”, which sometimes is referred to as a democracy of inclusive observation over the activities of ruling political class. At the same time, the mass media are increasingly interfacing with the political space in terms of the use of new information technologies in political communications. Secondly, the role of social media as a channel of political communication is growing. Within the framework of these media, social networks playing a growing role in the political mobilization of masses are becoming the most popular. Thereby, political actors obtain a possibility to coordinate their actions. Thirdly, advertising of political communications takes place. It is expressed in the subordination of political and communication interactions to the laws of advertising business concerning formation of political image of famous leaders. Advertising is supplemented by marketing of political communications, which consists in its impact on electorate by means of resources of “personal”, “software” and “information” influence in order to gain a place in the state and municipal power system. It is accompanied by distortion and fabrication of facts, tendentious and one-sided editing of misinfor-mation, discrediting of opponents with abnormal rhetoric, labelling and trolling. Three conditions described above constitute the context thanks to which, the “fake news” influence increasingly on political communications. The term “Fake news” is recognized as an expression of 2017 and refers to knowingly false, sensational information disseminated under the guise of news, which also includes the biased journalism, propaganda, hidden advertising, “post-telling” and trolling. Features of the “Fake news” communication technology are the following: high speed of distribution of such

messages, wider coverage of audience, difficulties of their identification. Different countries, de-spite the national, cultural and political differences, are trying to fight against the fake news both jointly and separately. In order to limit the impact of the “Fake news” on modern political communications, active countermeasures are being taken by social networks themselves (e.g. Facebook labels them as controversial). However, notable results have not yet been achieved. As a result, fakes have become a threat not only to the mass media and democracy as a whole, but also started to create “obstacles” for political and communication interaction between the real political actors. Researchers have to make additional efforts to analyse online reports in order to ensure the reliability of information disseminated and guarantee that it does not involve fake misinformation.

A sensitive stylistic approach to identify fake news on social networking:

Human inefficiency to distinguish between true and false facts poses fake news as a threat to logical truth, which deteriorates democracy, journalism, and credibility in governmental institutions. In this letter, we propose a computational-stylistic analysis based on natural language processing, efficiently applying machine learning algorithms to detect fake news in texts extracted from social media. The analysis considers news from Twitter, from which approximately 33,000 tweets were collected, assorted between real and proven false. In assessing the quality of detection, 86% accuracy, and 94% precision stand out even employing a dimensional reduction to one-sixth of the number of original features. Our approach introduces a minimum overhead, while it has the potential of providing a

high confidence index on discriminating fake from real news.

Optimal social trust path selection in complex social networks

Online social networks are becoming increasingly popular and are being used as the means for a variety of rich activities. This demands the evaluation of the trustworthiness between two unknown participants along a certain social trust path between them in the social network. However, there are usually many social trust paths between participants. Thus, a challenging problem is finding which social trust path is the optimal one that can yield the most trustworthy evaluation result. In this paper, we first present a new complex social network structure and a new concept of Quality of Trust (QoT) to illustrate the ability to guarantee a certain level of trustworthiness in trust evaluation. We then model the optimal social trust path selection as a Multi-Constrained Optimal Path (MCOP) selection problem which is NP-Complete. For solving this problem, we propose an efficient approximation algorithm MONTE K based on the Monte Carlo method. The results of our experiments conducted on a real dataset of social networks illustrate that our proposed algorithm significantly outperforms existing approaches in both efficiency and the quality of selected social trust paths.

Fake news detection regarding the Hong Kong events from tweets

The rapid development of network services has led to the exponential growth of online information and the increasing number of social media users. These services are exploited by malicious accounts that spread fake news and propaganda in vast user networks. Consequently, an automated solution for

fake news and deception detection is required. This paper introduces a new data set consisting of 2,366 tweets written in English, regarding the Hong Kong events (August, 2019), and a well-defined method for fake news detection that uses both linguistic and network features. Our approach is tested with experiments using 2 machine learning models, achieving high performance compared to previous research.

A survey on recent advances in machine learning techniques for fake news detection

Since the extent of ecological and societal issues, machine learning is particularly relevant in the perspective of fake messages in Social Media. Anyone can make a message and go viral which can be a fake or real one. Considering countries like India, the political parties during general elections do like this by spreading fake messages all over the country through social media groups. Some of the messages may be real but many are fake messages. Not only the fake messages containing false news but also spam messages can be spread through social media groups. This is a serious problem to the policy makers as well as the individual citizens. Researchers identify spam messages and fake news using various machine learning techniques. Deep Learning is a subset of machine learning which process like a human brain in solving different convoluted tasks. In fact, the society nowadays utilizes deep learning approaches because they can create successful text recognition issues, such as false news and spam identification, as such models operate on raw data efficiently through studying high-level features on their own. In this paper, various traditional machine learning methods and deep learning techniques that are used to identify fake and spam messages are

discussed. This paper also showcases a survey on different researches performed in fake news detection using traditional machine learning methods and Deep Neural Networks.

3. METHODOLOGY

Now-a-days all users are using social media to access news content but sometime some malicious users will alter genuine news and then spread fake news which may degrade social media fame and to avoid such fake news many existing algorithms were introduced but all those algorithms are based on traditional machine learning algorithms such as SVM or Random Forest. This algorithms lack of security and authorization.

The drawback of this system is algorithms lack of security and authorization.

To overcome this problem, we have proposed an integrated system for various aspects of blockchain and natural language processing (NLP) to apply machine learning techniques to detect fake news and better predict fake user accounts and posts. The Reinforcement Learning technique is applied for this process. To improve this platform in terms of security, the decentralized blockchain framework applied, which provides the outline of digital contents authority proof. More specifically, the concept of this system is developing a secure platform to predict and identify fake news in social media networks.

Benefits of this is Blockchain verify all hashcode before storing new block and if all blocks verification successful then only new block will store and this mechanism is called as proof of work or proof of authority. The concept of this system is developing a

secure platform to predict and identify fake news in social media networks.

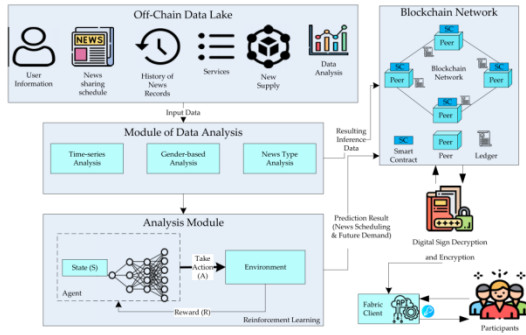


Fig 2 Proposed Architecture

Block Chain

A blockchain is a distributed ledger with growing lists of records (blocks) that are securely linked together via cryptographic hashes. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a Merkle tree, where data nodes are represented by leaves). The timestamp proves that the transaction data existed when the block was created. Since each block contains information about the previous block, they effectively form a chain (compare linked list data structure), with each additional block linking to the ones before it. Consequently, blockchain transactions are irreversible in that, once they are recorded, the data in any given block cannot be altered retroactively without altering all subsequent blocks. Blockchains are typically managed by a peer-to-peer (P2P) computer network for use as a public distributed ledger, where nodes collectively adhere to a consensus algorithm protocol to add and validate new transaction blocks. Although blockchain records are not unalterable, since blockchain forks are possible, blockchains may be considered secure by design and

exemplify a distributed computing system with high Byzantine fault tolerance. A blockchain was created by a person (or group of people) using the name (or pseudonym) Satoshi Nakamoto in 2008 to serve as the public distributed ledger for bitcoin cryptocurrency transactions, based on previous work by Stuart Haber, W. Scott Stornetta, and Dave Bayer. The implementation of the blockchain within bitcoin made it the first digital currency to solve the double-spending problem without the need of a trusted authority or central server. The bitcoin design has inspired other applications and blockchains that are readable by the public and are widely used by cryptocurrencies. The blockchain may be considered a type of payment rail. Private blockchains have been proposed for business use. Computerworld called the marketing of such privatized blockchains without a proper security model "snake oil"; however, others have argued that permissioned blockchains, if carefully designed, may be more decentralized and therefore more secure in practice than permissionless ones.

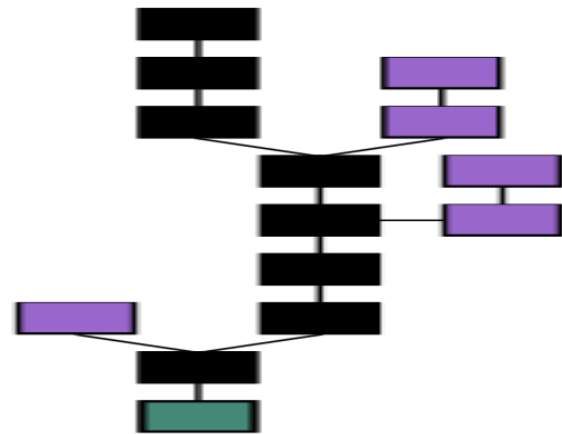


Fig 3 Block Chain formation

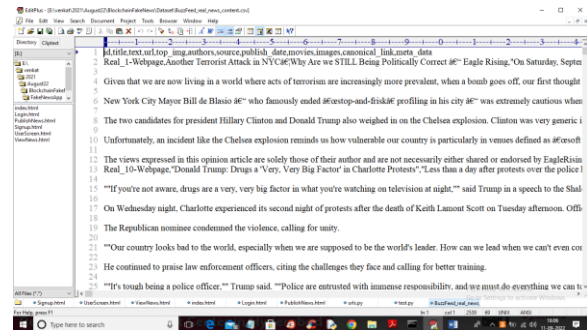
4. IMPLEMENTATION

Now-a-days all users are using social media to access news content but sometime some malicious users will alter genuine news and then spread fake news which may degrade social media fame and to avoid such fake news many existing algorithms were introduced but all those algorithms are based on traditional machine learning algorithms such as SVM or Random Forest. This algorithms lack of security and authorization.

In propose paper author is using Reinforcement Deep Learning algorithm to detect fake news and then applying Blockchain technology to provide security to NEWS data. Reinforcement algorithm will match similarity between old news and current news and if news contains more similarity then it will detect as Fake news and penalized the user and if news is REAL then algorithm will reward user and then store NEWS in Blockchain. Blockchain store each news as block/transaction and associate each block with unique hashcode and then store this block in multiple nodes as Decentralized storage and if one node is down then Blockchain will access News from other working nodes. Blockchain verify all hashcode before storing new block and if all blocks verification successful then only new block will store and this mechanism is called as proof of work or proof of authority.

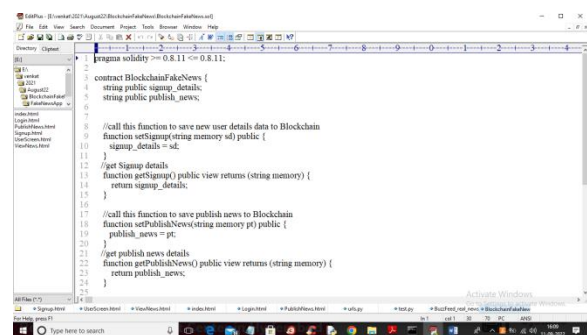
To detect fake news in propose paper applying NLTK (natural language tool kit) technique to remove STOP WORDS, special symbols, data normalization, converting words into numeric vector (numeric vector will assign average count of each word), STEMMING and lemmatization. Vector will be input to Reinforcement algorithm to train FAKE news model and this trained model will be applied on all TEST news to predict weather news is fake or real.

Reinforcement will take each INPUT as state and then apply model to predict ACTION and if action is Fake then algorithm will penalized user and if action is real then user or news publisher will get reward. To implement this project author has used BUZZ NEWS and many more dataset but we are using BUZZ news dataset.



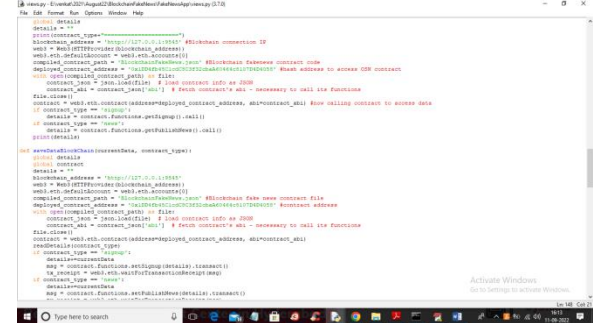
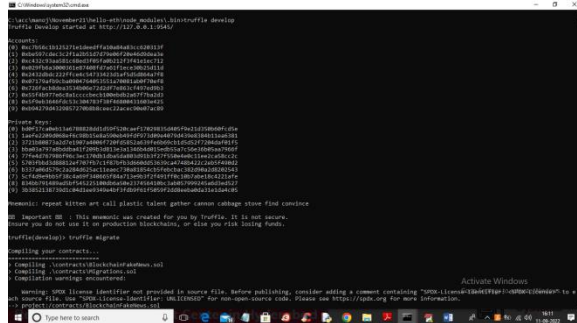
In above we can see first row contains dataset column names and remaining rows contains dataset values and by using above dataset we will train Reinforcement algorithm.

To store or access data from Blockchain we need to develop SMART CONTRACT which will contains functions to STORE and READ data.



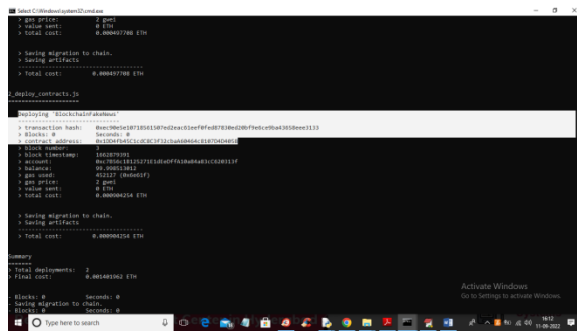
In above fig we have define smart contract functions to store USER & NEWS details and we need to deploy this contract in Blockchain server and for deployment we need to follow below steps

First go inside 'hello-eth/node_modules/.bin' folder and then double click on 'runBlockchain.bat' file to get below screen



In above we can see read red colour comments to know about how to call Blockchain function to store and read data using Python program.

In above we can see Blockchain generated Private keys and default account address and now type command as 'truffle migrate' and then press enter key to deploy contract and get below output



5. EXPERIMENTAL RESULTS



Fig 4 Home Page

In above fig in white colour text we can see Blockchain Fake News contract deployed and we got contract address also and this address we need to specify this address in python program to access above contract to store and read data from Blockchain.

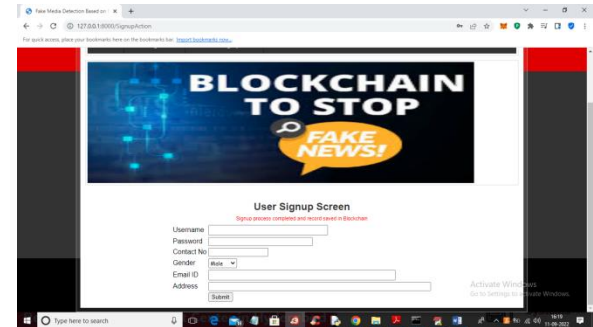


Fig 5 Signup Page



Fig 6 Login Page



Fig 7 Main Page



Fig 8 Fake & real news graph

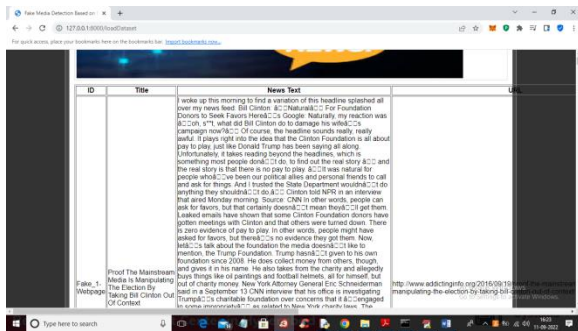


Fig 9 Upload News

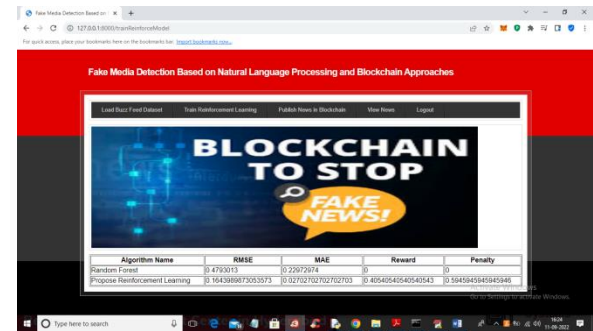


Fig 10 Reinforcement training output

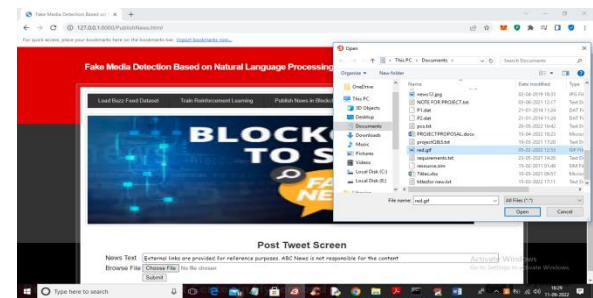


Fig 11 upload news

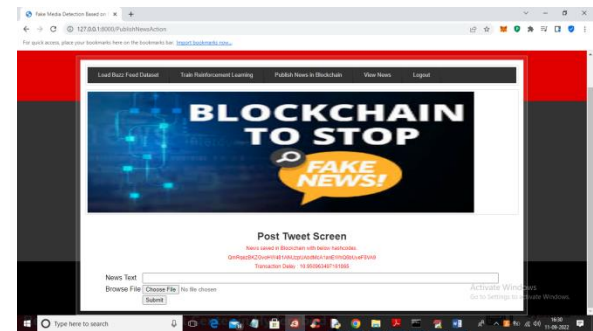


Fig 12 News saved in blockchain hashcode

Fig 13 Prediction result that stored in blockchain
hashcode

6. CONCLUSION

Fake news sharing is one of the popular research problems in recent technology based on lack of security and trust in terms of the truth of shared news in social media. In this article, we have presented the combination of blockchain and machine learning techniques to provide solutions and design a trust-based architecture toward shared news online. We have applied the reinforcement learning technique, a learning based algorithm, to make a strong decision-making architecture and combine it with blockchain framework, smart contract, and customized consensus algorithm, which is well fit for the Proof-of-Authority protocol. Social media plays a key role in this process. The shared information platform contains fake news, and its a beneficial challenge to enhance and investigate the Proof-of-Authority protocol and user validation.

REFERENCES

- [1] V. P. Miletskiy, D. N. Cherezov, and E. V. Strogetskaya, "Transformations of professional political communications in the digital society (by the example of the fake news communication strategy)," in Proc. Commun. Strategies Digit. Soc. Workshop (ComSDS), 2019, pp. 121–124.
- [2] N. R. de Oliveira, D. S. V. Medeiros, and D. M. F. Mattos, "A sensitive stylistic approach to identify fake news on social networking," IEEE Signal Process. Lett., vol. 27, pp. 1250–1254, 2020.
- [3] G. Liu, Y. Wang, and M. Orgun, "Optimal social trust path selection in complex social networks," in

Proc. AAAI Conf. Artif. Intell., vol. 24, 2010, pp. 1391–1398.

- [4] M. N. Nikiforos, S. Vergis, A. Styliadou, N. Augoustis, K. L. Kermanidis, and M. Maragoudakis, "Fake news detection regarding the Hong Kong events from tweets," in Proc. Int. Conf. Artif. Intell. Appl. Innov. Greece: Springer, 2020, pp. 177–186.
- [5] A. R. Merryton and G. Augasta, "A survey on recent advances in machine learning techniques for fake news detection," Test Eng. Manag, vol. 83, pp. 11572–11582, 2020.
- [6] X. Dong, U. Victor, S. Chowdhury, and L. Qian, "Deep two-path semisupervised learning for fake news detection," 2019, arXiv:1906.05659. [Online]. Available: <http://arxiv.org/abs/1906.05659>
- [7] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," Trans. Emerg. Telecommun. Technol., vol. 31, no. 2, p. e3767, Feb. 2020.
- [8] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bidirectional LSTM-recurrent neural network," Procedia Comput. Sci., vol. 165, pp. 74–82, Jan. 2019.
- [9] 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.
- [10] M. Mahyoob, J. Algaraady, and M. Alrahaili, "Linguistic-based detection of fake news in social media," Int. J. English Linguistics, vol. 11, no. 1, p. 99, Nov. 2020.



[11] A. Koirala, “COVID-19 fake news classification using deep learning,” Tech. Rep., 2020. [Online]. Available: <https://www.cs.ait.ac.th/xmlui/handle/123456789/981>

[12] H. Gill and H. Rojas, “Chatting in a mobile chamber: Effects of instant messenger use on tolerance toward political misinformation among south Koreans,” *Asian J. Commun.*, vol. 30, no. 6, pp. 470–493, Nov. 2020.

[13] J. L. Alves, L. Weitzel, P. Quaresma, C. E. Cardoso, and L. Cunha, “Brazilian presidential elections in the era of misinformation: A machine learning approach to analyse fake news,” in *Proc. Iberoamerican Congr. Pattern Recognit.* Havana, Cuba: Springer, 2019, pp. 72–84.

[14] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de Medeiros, and D. M. F. Mattos, “Identifying fake news on social networks based on natural language processing: Trends and challenges,” *Information*, vol. 12, no. 1, p. 38, Jan. 2021.

[15] D. Mouratidis, M. N. Nikiforos, and K. L. Kermanidis, “Deep learning for fake news detection in a pairwise textual input schema,” *Computation*, vol. 9, no. 2, p. 20, Feb. 2021