World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:8, 2024

Detecting Fake News: A Natural Language Processing, Reinforcement Learning, and Blockchain Approach

Ashly Joseph, Jithu Paulose

Abstract**—**In an era where misleading information may quickly circulate on digital news channels, it is crucial to have efficient and trustworthy methods to detect and reduce the impact of misinformation. This research proposes an innovative framework that combines Natural Language Processing (NLP), Reinforcement Learning (RL), and Blockchain technologies to precisely detect and minimize the spread of false information in news articles on social media. The framework starts by gathering a variety of news items from different social media sites and performing preprocessing on the data to ensure its quality and uniformity. NLP methods are utilized to extract complete linguistic and semantic characteristics, effectively capturing the subtleties and contextual aspects of the language used. These features are utilized as input for a RL model. This model acquires the most effective tactics for detecting and mitigating the impact of false material by modeling the intricate dynamics of user engagements and incentives on social media platforms. The integration of blockchain technology establishes a decentralized and transparent method for storing and verifying the accuracy of information. The Blockchain component guarantees the unchangeability and safety of verified news records, while encouraging user engagement for detecting and fighting false information through an incentive system based on tokens. The suggested framework seeks to provide a thorough and resilient solution to the problems presented by misinformation in social media articles.

Keywords**—**Natural Language Processing, Reinforcement Learning, Blockchain, fake news mitigation, misinformation detection.

I. Introduction

IN the era of digital information, social media platforms have become a primary source of news usage for millions of users worldwide. The ease of sharing and the rapid dissemination of information on these platforms have led to an unprecedented spread of false information, often in the form of fake news articles. The proliferation of fake news has been exacerbated by the anonymity and lack of accountability in social media, as well as the potential for malicious actors to manipulate public opinion and spread conflicts [1]. The consequences of this phenomenon are far-reaching, ranging from political polarization and erosion of trust in institutions to public health crises and financial market instability. As the volume and sophistication of fake news continue to grow, it is crucial to develop effective strategies to identify and mitigate the spread of false information in social media news articles [3].

Ashly Joseph is researcher/engineer from California, USA (e-mail: ashlyelsy@gmail.com)

A. Importance of Identifying and Reducing Fake News

Detection and reduction of misleading information in news pieces on social media are crucial for several reasons. To begin with, false information may have major negative impacts on both people and society at large. It has the potential to mislead individuals into making incorrect choices, particularly in areas such as healthcare, money, and politics. Furthermore, the spread of inaccurate information has the potential to weaken democratic procedures by establishing an impact on public opinion and electoral results. Additionally, it can worsen inequality in society and contribute to an atmosphere of fear and mistrust [2]. Moreover, the spreading of false information can result in adverse economic effects, including financial loss for enterprises and investors that depend on precise data for their decision-making processes. Hence, it is vital to create efficient strategies for detecting and countering misinformation in order to safeguard people, preserve peace in society, and preserve the trustworthiness of democratic institutions and markets.

B. Constraints of Current Approaches

The existing methods for detecting and reducing wrong information in news stories on social media have several restrictions. Standard verification techniques, which require journalists and specialists to manually verify statements, are both time-consuming and require a significant number of resources. They frequently have difficulties in keeping up with the extensive quantity and rapid spread of false information on social media platforms [8]. Automated methodologies, such as those utilizing machine learning and NLP, have demonstrated potential in effectively identifying fraudulent news with more efficiency. Nevertheless, these approaches frequently depend on superficial characteristics and can face difficulties in capturing the complex details and situations of language that are crucial for distinguishing between genuine and false material. In addition, several current methods concentrate exclusively on the substance of news stories and fail to take into account the wider social and behavioral factors that contribute to the spread of false information, such as user interactions and network effects. Consequently, there is a want for more extensive and advanced methods that can accurately detect and reduce erroneous information in social media news items [3].

II. Algorithm Details

A. Natural Language Processing

The area of study called NLP looks at how computers and human language interact. The goal is to make computers able to understand, analyze, and create meaningful insights from human language [15]. NLP is a rapidly evolving discipline in software engineering that aims to provide computers with the ability to read and interpret human language in a way that resembles human understanding [4].

NLP combines computational models of human language based on linguistic rules with data, artificial intelligence (AI), and advanced learning algorithms. This combination enables computers to fully grasp human language in both written and spoken form, accurately capturing the intentions and viewpoints of the speaker or writer. NLP has a wide range of practical applications that are numerous and vast. It is of extreme significance in the advancement of systems that simplify the translation of text across different languages, respond to spoken instructions, and efficiently and properly summarize extensive amounts of material.

NLP technologies provide the basis for voice-activated GPS systems, digital assistants, speech-to-text transcription programs, customer care chatbots, and other online purchasing solutions that prioritize ease. In addition to its use in consumer applications, NLP plays a crucial role in improving important corporate processes, increasing employee productivity, and optimizing operations on a broad scale. Enterprises utilize NLP to examine consumer feedback, automate repetitive operations, enhance communication channels, and enable more informed decision-making by extracting practical insights from unorganized data. NLP plays a crucial role in modern software applications that require seamless and intelligent interaction with human language by bridging the gap between human communication and computer translation [5].

B. Reinforcement Learning Model

RL is a separate method that falls in between supervised and unsupervised learning. In RL, an agent learns to make optimal decisions by taking actions in an environment and receiving feedback in the form of rewards for favorable actions and penalties for unfavorable ones. The environment offers rewards for favorable actions and penalties for unfavorable ones, gradually leading the agent towards achieving optimal behavior. The process starts with the agent initiating an activity within the environment. The environment provides feedback, indicating whether the activity was helpful (rewarding) or harmful (punishing). The feedback loop enables the agent to acquire knowledge from its conversations, enhancing its approach to optimize the total rewards obtained over time [6]

RL, in contrast to supervised learning, adapts its behavior by communicating with the environment rather than relying on labeled training data. This makes it especially suited for tasks that involve sequential decision-making and where outcomes are impacted by these decisions. Regarding the identification of fake news, a RL agent may be trained to recognize patterns that differentiate between genuine and fabricated news pieces. Through iterative adjustments of its strategies in response to received incentives and penalties, the agent enhances its capacity to effectively categorize novel articles, hence strengthening the durability of the detection system.

C. Blockchain

Blockchain technology is an innovative approach to data storage that effectively prevents change, unauthorized access, and fraudulent activities. A blockchain is a decentralized and duplicated digital ledger that records and distributes transactions across a network of computers. Each instance of a new transaction on the blockchain is documented and duplicated on the ledger of every participant, guaranteeing both transparency and security. Each individual unit in the blockchain is composed of several transactions, and these units are interconnected in a sequential chain. Distributed Ledger Technology (DLT) refers to a decentralized database that is operated and maintained by various contributors [6], [8]

Blockchain records are unchangeable and verified using cryptography, offering a chronologically recorded history of all transactions. This functionality enables any authorized user to retrospectively track the whole transaction history, hence augmenting accountability and transparency. Blockchain technology has uses that go beyond just cryptocurrencies. It is employed across several industries, including banking, supply chain management, healthcare, and real estate, to bolster security, mitigate fraud, and promote operational efficiency. Blockchain technology is facilitating safe and transparent recording and verification of transactions, hence fostering trust-based digital interactions and innovations.

III. Literature Review

The rapid growth of communication technology and the widespread use of smart devices have led to a significant increase in data traffic, resulting in the creation of enormous volumes of data every second [2].

The concept of big data has received considerable attention in the last ten years because of its capacity to provide essential insights and advantages, such as lower expenses, accelerated decision-making, and advancements in product creation across diverse sectors. Traditional data analysis approaches are typically insufficient for addressing the complexity and scope of big data due to the continuous nature of large data sources [4].

NLP methods have been widely used to examine the textual content of news stories and identify significant characteristics for the identification of false news. Only one study provides an in-depth look of NLP-based techniques, including speech analysis and network-based methodologies [8]. Their findings emphasize the efficacy of combining vocabulary, syntactic, and semantic characteristics to distinguish between authentic and fake news pieces [7]. A different research utilizes language characteristics such as lexical diversity, readability measures, and emotional sentiment to identify misleading information in news stories [9]. Their work shows the capacity of NLP approaches to capture the nuanced distinctions between genuine and fake news.

The increasing number of false news has emerged as an important issue in the current digital environment, mostly attributed to the extensive use of social media platforms [5]. Verifying the genuineness of information shared on these platforms is of utmost importance, however it can be difficult. The Fake News and Fake URL Detection Using Bi-CNN (FNU-BiCNN) model is a suggested approach that uses Natural Language Toolkit (NLTK) features, such as stop words and stem words, to preprocess the data [13]. The model uses Long Short-Term Memory (LSTM), batch normalization, and dense layers to compute Term Frequency-Inverse Document Frequency (TF-IDF) [9]. The datasets are trained using Bi-LSTM with AutoRegressive Integrated Moving Average (ARIMA) and Convolutional Neural Networks (CNN), and several machine learning techniques are employed for classification. This performance surpassed that of other machine learning methods, including Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), and Naive Bayes. Accuracy, recall, and F1-Score were used to evaluate performance and effectiveness [8]

Machine learning algorithms have been extensively used to categorize news items according to the extracted characteristics. A research study suggests a hybrid model that integrates a Recurrent Neural Network (RNN) with a CNN to effectively capture both the temporal and contextual information present in articles. Their methodology attains a notable level of precision in identifying fabricated news by relying on the advantages of deep learning frameworks [10]. Different research provides an in-depth review of machine learning-based strategies for detecting false news. It explores several methods of feature engineering, classification algorithms, and evaluation measurements. The authors highlight the significance of taking into account many factors, including content, user involvement, and network structure, in order to enhance the effectiveness of false news detection algorithms [8].

Another important part of data and software engineering is the concept of report grouping, which involves carefully arranging records into distinct groups [9]. Due to the ongoing progress in personal computers and technology, there has been a consistent rise in the number of reports. This calls for the need to effectively classify them based on their content. Text classification is a widely used technique to group text into distinct categories. This process involves numerous steps that may be tackled using various approaches [10]. The choice of suitable techniques for each category is crucial in improving the effectiveness of text processing. Classifying archives according to their content is an intricate task that plays a crucial role in the endeavors of data specialists and researchers. It is essential for applications in the fields of design, organization, and effective handling of vast amounts of information. This is especially crucial for publishers, publications, bloggers, and organizations that handle huge amounts of content [11].

IV. Proposed Framework

The framework that we suggest combines NLP, RL, and Blockchain technologies to develop a complete robust solution for detecting and reducing false content in social media articles. The framework begins by gathering a variety of material from different social media sites and performing preprocessing on the data to ensure its quality and consistency. NLP techniques are used to extract complete linguistic and semantic features from the news stories, effectively capturing the subtle nuances and contextual information given by the any language. The features are used as input for a RL model. This model learns the best strategies for detecting and reducing misleading information by modeling the complex nature of user interactions and objectives on social media sites [12].

The RL model is structured as a Markov Decision Process (MDP) and trained using a carefully designed reward function that encourages the detection and reduction of false information [1]. Ultimately, the integration of Blockchain technology into the framework establishes a decentralized and transparent system specifically designed for storing and validating the accuracy of news chunks. The Blockchain component guarantees the unchangeability and safety of the verified news records, while motivating users to participate in identifying and reducing fake information by using a token-based incentive system. Our framework utilizes advanced technology to effectively address the complex issue of misinformation in social media reports [3].

A. Data Collection and Preprocessing

In order to build and evaluate our framework, we will gather a large diverse dataset of news stories collected from widely used social media sites like Twitter, Facebook, and Reddit. The collection will consist of a mix of genuine and fabricated news stories covering a wide range of subjects, including politics, health, science, and entertainment. We will utilize the APIs offered by these platforms to collect significant news items and their corresponding metadata, including user engagements (such as likes, shares, and comments), timestamps, and user profiles. In order to guarantee the dataset's quality and acceptance, we will utilize stratified sampling methods to choose articles from various time periods, geographic locations, and user demographics. In addition, we will work together with fact-checking organizations and topic specialists to acquire accurate labels for a portion of the gathered articles. These labels will be utilized to train and validate our models. The dataset will be regularly updated to accurately represent the ever-changing nature of social media news and to capture developing trends and patterns in the spread of incorrect information [11].

Before starting feature extraction and building a model, the obtained dataset will undergo extensive cleaning and normalization to guarantee the quality and consistency of the data. Our first step will be eliminating duplicate articles and those that contain missing or incomplete information. Subsequently, we will execute text preparation operations, such as eliminating HTML elements, special characters, and URLs, in order to get purified and standardized text data. In addition, we will process multilingual articles by identifying the language of each article and using suitable language-specific preprocessing procedures. In order to reduce the effects of noise

and outliers, we will utilize statistical methods such as z-score normalization on the numerical metadata linked to the articles, such as user interaction counts. In addition, we will implement data anonymization techniques to safeguard user privacy. This involves either deleting or encrypting personally identifying information (PII) from the dataset. The cleaned and standardized dataset will be saved in a safe and scalable database system, such as MongoDB or Cassandra, to enable efficient retrieval and processing in the coming phases of the framework [5].

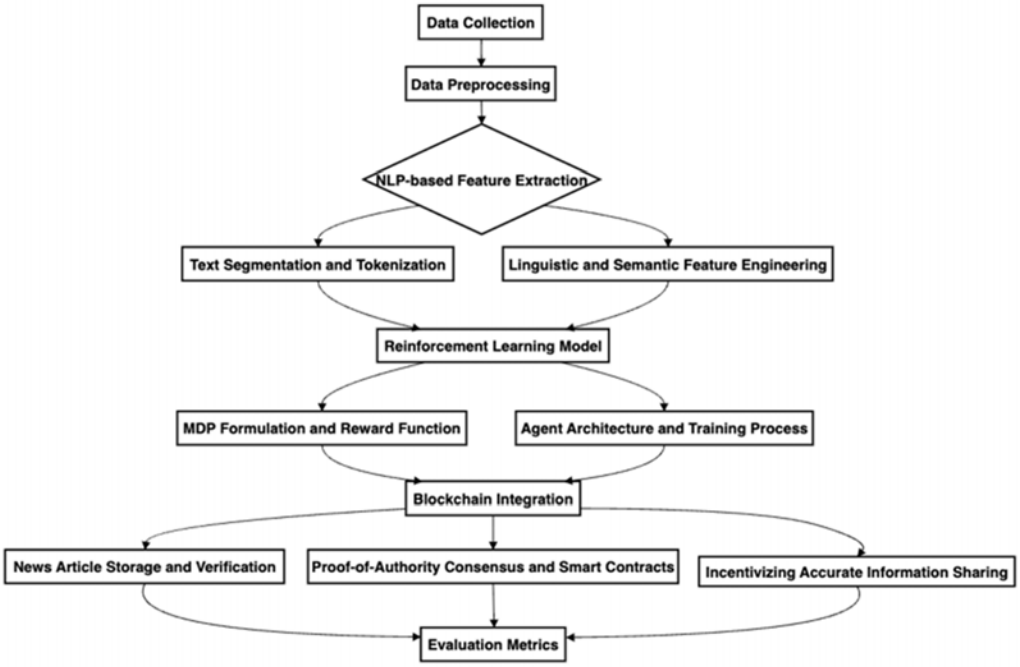

Fig. 1 Proposed architecture framework

B. NLP-Based Feature Extraction

In order to get meaningful linguistic and semantic characteristics from the news articles, our initial step will include performing text segmentation and tokenization. Text segmentation is the process of dividing articles into smaller components, such as paragraphs, sentences, or phrases, based on the structure and syntax of the text. This stage helps in capturing the particular context and linguistic features included in the articles [6]. Subsequently, we will employ tokenization methodologies to divide the segmented text into individual words or tokens. We will conduct experiments using several tokenization methods, including word-based, character-based, and subword-based tokenization, in order to determine the most efficient strategy for our dataset. In addition, we will address typical difficulties encountered in tokenization, such as managing punctuation, contractions, and named entities, in order to guarantee the high standard and consistency of the tokenized text. The segmented and tokenized text will be used as a framework for the next feature [14].

Building upon the segmented and tokenized text, we will create an extensive set of linguistic and semantic features to capture the complexity and context of the language used in the news items. Initially, we will extract basic linguistic characteristics, such as the frequencies of words, part-of-speech tags, and named entities, in order to capture the syntactic and grammatical attributes of the text. Subsequently, we will use advanced NLP approaches, such as dependency parsing and semantic role labeling, to determine the connections between words and their respective functions in the sentences. In order to maintain the semantic meaning and context of the text, we will utilize techniques such as word embeddings (e.g., Word2Vec, GloVe). These approaches represent words as dense vectors in a high-dimensional space, therefore remaining their semantic relationships [8].

We will also look into contextualized word embeddings, such as BERT (Bidirectional Encoder Representations from Transformers) and Elmo (Embeddings from Language Models), that provide dynamic word representations based on the surrounding context. In addition, we will utilize lexicon-based and machine learning methods to extract sentiment and emotion characteristics from the articles, in order to capture the affective aspects of the language utilized. Additional possible elements include readability ratings, stylometric characteristics, and domain-specific knowledge graphs. The features that have been retrieved will be merged together to create a feature vector representation with a large number of dimensions for each article. This representation will then be utilized as input for the RL model.

C. Reinforcement Learning Integration

The issue of identifying and removing false information on social media can be represented as an MDP. The MDP will consist of states that represent the current status of a news article, actions which correspond to decisions made by the RL agent, transitions that model the probability of moving between

states, and a reward function that is designed to promote accurate identification and reduction of false information. The incentive function will be carefully designed to correspond with the aims of the framework, offering positive rewards for accurately identifying counterfeit articles and diminishing their prominence, and negative rewards for misclassifying or enabling the dissemination of fraudulent material. The MDP formulation allows the RL agent to develop optimal policies that maximize the predicted cumulative rewards over time [6].

The RL agent will utilize a deep learning framework to efficiently acquire knowledge and represent complex patterns in the news articles' high-dimensional features. We will conduct experiments using several neural network designs, including CNNs, RNNs, and Graph Neural Networks (GNNs), in order to capture dependencies between spatial, time, and structural features. The agent will undergo training utilizing a mix of offline and online learning methodologies. The offline training will consist of utilizing supervised learning, imitation learning, and value-based RL techniques on a tagged sample of news stories. Online learning will utilize policy gradient techniques and exploration tactics to continuously modify and adjust the agent's policy using real-time feedback and rewards. The use of techniques such as transfer learning, domain adaptation, and normalization will be implemented to guarantee the adaptability and uniformity of the system. Additionally, safety and fairness requirements will be integrated to remove any potential biases [13].

D. Blockchain Integration

We will integrate Blockchain technology to create an immutable and easily verifiable system for storing and validating the accuracy of news articles. Articles that have been verified will be kept on the Blockchain as unchangeable records. These records will be represented as unique digital assets with cryptographic hashes and timestamps. Smart contracts will automate the process of storing and retrieving content, by establishing verification criteria such as mandatory fact-checking confirmations or consensus among validators. Methods like as sharding, off-chain processing, and state channels will ensure scalability and effectiveness, while privacy-preserving technologies will protect sensitive data and ensure user privacy.

We will use a Proof-of-Authority (PoA) consensus method, wherein reliable validators are tasked with validating transactions and creating new blocks on the Blockchain. Smart contracts will manage the process of selecting, rotating, and holding validators responsible. They will establish the criteria and requirements for being a validator and create tools to identify and punish harmful actions. The utilization of Byzantine Fault Tolerance (BFT) algorithms and threshold cryptography techniques would guarantee the ability of the Proof of Authority (PoA) consensus to withstand and recover from failures and errors. The integration of PoA consensus mechanism with smart contracts would establish a robust, effective, and open framework for verifying and storing authenticated news items on the Blockchain [3].

In order to encourage user involvement in the identification and reduction of incorrect information, we will develop a token-based incentive system. Users will get native cryptocurrency tokens as rewards for identifying and reporting fake news, presenting proof for fact-checking, or serving as validators. The provision of token incentives will ensure that users' interests are in line with the goals of the framework, therefore encouraging the exchange of accurate information and discouraging the spread of fraudulent information. Systems will be developed to generate, distribute, and circulate tokens in order to guarantee long-term viability and worth. By using Blockchain technology, token transactions will be marked by transparency, security, and immutability. This will guarantee the fairness and integrity of the incentive system [5], [8].

E. Assessment Criteria

In order to evaluate the effectiveness and performance of our suggested framework, we will establish and execute a complete range of assessment criteria. The measurements will include several facets of the framework, such as the precision in detecting false information, the efficiency of the RL model, the impact of Blockchain integration, and the overall user satisfaction [11], [8].

- *Detection Accuracy:* The accuracy of our system in detecting incorrect data will be evaluated using standard categorization measures, including precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC) curve. The metrics will be calculated by comparing the projected labels with the actual labels of the news items in our dataset. In addition, we will assess the framework's capacity to address various forms of misinformation, including entirely invented stories, deceptive material, and altered media.
- *RL Model Efficiency:* The efficiency of our RL model will be evaluated by monitoring variables such as convergence time, sample complexity, and computational overhead. Convergence time refers to the speed at which the RL agent acquires an optimum policy for detecting and reducing the impact of erroneous information. Sample complexity is the quantification of the number of interactions an agent needs with the environment in order to reach a specific level of performance. Computational overhead refers to the amount of time and resources used by the RL model for both training and inference.
- *Blockchain Performance:* We will assess the efficiency of our Blockchain integration by measuring parameters such as transaction throughput, latency, and scalability. Transaction throughput refers to the capacity of the Blockchain network to complete a certain number of transactions, such as verifying news articles, within a given time frame, often measured in seconds. Latency is the duration it takes for a transaction to be verified and included on the Blockchain. Scalability evaluates the capacity of the Blockchain network to manage growing workloads and sustain performance as the quantity of users and transactions expands.
- *User Engagement and Satisfaction:* In order to assess the influence of our framework on user experience, we will

carry out user tests and surveys to collect input on metrics such as usability, confidence in the system, and perceived worth of the token incentives. In addition, we will monitor user engagement metrics, including the quantity of active users, the frequency of interactions, and the caliber of user contributions, such as the accuracy of flagging and involvement in fact-checking.

- *Robustness and Generalization:* Robustness and generality will be assessed by evaluating the performance of our system across various datasets, social media platforms, and topics. To evaluate the framework's capability to manage differences in data distribution, user behavior, and information contexts, we will employ approaches such as cross-validation, domain adaptation, and transfer learning.

V. Conclusion

Identifying fake news is a critical effort in the present era of digital technology, when the spread of false data can result in significant consequences. The proposed system combines NLP, RL, and Blockchain technologies to successfully tackle the complex task of identifying and limiting the spread of wrong information in social media articles. The system provides a complete and resilient solution by utilizing modern NLP techniques to extract features, employing a RL model for dynamic decision-making, and integrating Blockchain for safe and transparent verification. The framework's effectiveness will be evaluated utilizing a varied range of evaluation criteria, such as detection precision, efficiency of the RL model, performance of the Blockchain, user involvement and satisfaction, as well as robustness and generalization. These measurements will offer an extensive viewpoint on the system's performance and its effectiveness in combating misinformation on social media. Implementing this approach effectively could greatly reduce the spread of false information, increase user confidence in social media platforms, and create a more knowledgeable and honest online environment. This research makes a valuable contribution to the current global efforts to address disinformation by using advanced technology and a comprehensive technique. Its goal is to build a digital information environment that is safer and more reliable for people throughout the world.

References

[1] Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., et al. 2020. *Weak Supervision for Fake News Detection via Reinforcement Learning.* Proceedings of the AI Conference on Artificial Intelligence, 34(01): 516–523.

[2] Conroy, N. K., Rubin, V. L., & Chen, Y. 2015. *Automatic deception detection: Methods for finding fake news*. Proceedings of the Association for Information Science and Technology, 52(1): 1–4.

[3] Yi-Chin Chen, Zhao-Yang Liu, and Hung-Yu Kao. 2017. *IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification.* In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 465–469, Vancouver, Canada. Association for Computational Linguistics.

[4] Popat, K., Mukherjee, S., Yates, A., & Weikum, G. 2018. *DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning*. arXiv https://doi.org/10.48550/arxiv.1809.06416.

[5] Xing, J., Wang, S., Zhang, X., & Ding, Y. 2021. HMBI: *A New Hybrid Deep Model Based on Behavior Information for Fake News Detection*. Wireless Communications and Mobile Computing, 2021: 1–7.

[6] Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A., & Krause, A. 2018. *Fake News Detection in Social Networks via Crowd Signals*. https://doi.org/10.1145/3184558.3188722.

[7] Elhadad, Mohamed K. et al. "*Fake News Detection on Social Media: A Systematic Survey*." 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM) (2019): 1-8.

[8] Van Der Walt, E., & Eloff, J. 2018. *Using Machine Learning to Detect Fake Identities: Bots vs Humans*. IEEE Access, 6: 6540–6549.

[9] Alkhodair, S. A., Ding, S. H. H., Fung, B. C. M., & Liu, J. 2020. *Detecting breaking news rumors of emerging topics in social media*. Information Processing & Management, 57(2): 102018.

[10] Ni, S., Li, J., & Kao, H.-Y. 2021. *MVAN: Multi-View Attention Networks for Fake News Detection on Social Media*. IEEE Access, 9: 106907–106917.

[11] Amer E, Kwak K-S, El-Sappagh S. *Context-Based Fake News Detection Model Relying on Deep Learning Models.* Electronics. 2022; 11(8):1255. https://doi.org/10.3390/electronics11081255

[12] Girgis, S., Amer, E., & Gadallah, M.E. (2018). *Deep Learning Algorithms for Detecting Fake News in Online Text.* 2018 13th International Conference on Computer Engineering and Systems (ICCES), 93-97.

[13] Zhou, X., & Zafarani, R. 2019b. *Network-based Fake News Detection: A Pattern-driven Approach*. arXiv. https://doi.org/10.48550/arxiv.1906.04210

[14] Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F. *The spread of low-credibility content by social bots.* Nat Commun. 2018 Nov 20;9(1):4787. doi: 10.1038/s41467-018-06930-7. PMID: 30459415; PMCID: PMC6246561.

[15] Sadiku M.N.O., Zhou Y., Musa S.M. *Natural Language Processing.* Int. J. Adv. Sci. Res. Eng. 2018;4:68–70. doi: 10.31695/ijasre.2018.32708.