International Journal of
ADVANCED RESEARCH & INNOVATIONS

Volume 15, Issue-7
July, 2025, P05
ISSN: 2319-9253

FAST TEXT EMBEDDINGS AND DEEP LEARNING FOR ROBUST DETECTION OF DEEPFAKES IN SOCIAL MEDIA TWEETS

[1]Dr T.Veeranna, [2]Sd.Iliyaz Ali, [3]Ch.Udaykoteswara Rao,
[4]D.Harsha Vardhan Reddy, [5]M. Gayathri
[1]Associate Professor, Dept. of CSE (AI&ML), Sai Spurthi Institute of Technology, Khammam, Telangana, India.
[2,3,4,5]B.TechStudents, Dept. of CSE(AI&ML), Sai Spurthi Institute of Technology, Khammam, Telangana, India.

Abstract: The increasing prevalence of deep fake technology has prompted apprehension regarding the dissemination of inaccurate information on social media. This paper illustrates a deep learning-based approach to identifying deep fake tweets, particularly those generated by machines. This will mitigate the detrimental effects of false information on the internet. Our approach categorizes tweets into categories by employing Fast Text embeddings and deep learning models. We employ Fast Text embeddings to generate dense vector models after preprocessing the tweet text. The distinction between genuine and fraudulent tweets is determined by the semantic information regarding tweet topics that these embeddings accumulate. We incorporate these embeddings into a deep learning model, such as a CNN or a Long Short-Term Memory (LSTM) network, to determine whether the tweets are genuine or fabricated. Machine-generated tweets are generated using contemporary text generation algorithms that have been instructed on a collection of tagged tweets. Research conducted on a real-world tweet collection demonstrates that our methodology is effective in identifying tweets that were generated by algorithms. Our approach is significantly more precise than other methods for identifying social media deep fakes. In general, our proposed approach is a dependable and efficient approach to identify tweets that were generated by machines and to halt the dissemination of inaccurate information on social media.

Keywords: Deep fake detection, deep learning, Fast Text embeddings, machine-generated tweets.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are properly cited.

1. Introduction

Deep fake technology is responsible for the massive volume of false and erroneous information that circulates on social media platforms, which is a major issue. Deep fakes are artificial intelligence-generated images, movies, or sounds that depict fraudulent events or persons saying things they never uttered. They jeopardize the integrity of online materials. Tweets, unlike other kinds of digital communication, are particularly easy to update since they are so brief and simple to send. This research proposes a novel method for detecting tweets created by computers, particularly those generated by deep fake algorithms. It employs deep learning algorithms. The proposal is intended to address these issues. To distinguish between actual and false tweets, we use a combination of recent deep learning models and Fast Text embeddings to improve text representation. We increased the capacity to distinguish between objects that need to be classified by leveraging the semantic complexity of Fast Text embeddings. These embeddings convert contextual and grammatical information from Twitter sentences into dense vector representations. Before we convert Twitter texts into Fast Text embeddings, we need to ensure that they are consistent and easy to read.

These embeddings serve as input for a powerful classification model, such as a CNN or LSTM network, which can distinguish between authentic and false tweets. To train and evaluate our system, we employ a tagged dataset of tweets created with new text creation algorithms that are similar to how machine-generated content appears in the real world. Our suggested method works fairly well for detecting tweets created by machines, as evidenced by our examination of a large set of real tweets. The results suggest that our method of detecting deep fakes on social networking sites outperforms those

already available. Our system successfully distinguishes between original and altered content, significantly reducing the impact of bogus information on the internet. As a result, information shared on social media platforms is more likely to be factual and accurate. Using Fast Text embeddings and deep learning, this study finally presents a strong solution to the important challenge of identifying tweets created by computers.

2. Literature Survey

Zhang, H., Li, Y., & Chen, M. (2024This study uses deep learning models and Fast Text embeddings to identify machine-written messages on social media platforms. Fast Text embeddings successfully capture sub word-level semantics, allowing a complete linguistic description of tweets. The hybrid neural network design employed in this study includes recurrent layers for processing sequences and convolution layers for feature extraction. The results demonstrate significant improvements in data processing speed and accuracy when compared to older methods. The approach is particularly effective at detecting tiny trends in multilingual datasets. This makes it a scalable tool to monitor social media.

Gao, X., Wu, P., & Liu, Z. (2024)This study investigates how successfully Fast Text embeddings detect Twitter deep fake messages. Because it employs Fast Text's out-of-vocabulary word handling and sub word unit representation, the proposed method excels at distinguishing between machine-generated material and real text. The researchers combined these embeddings into a neural network design that achieves a decent balance of accuracy and computing speed. This makes them more accessible to a wider range of users. The testing results reveal that the model performs effectively on a variety of datasets, including those containing specially produced tweets created by computers. The authors emphasize how well the strategy prevents automated campaigns from disseminating incorrect information and discuss how it could be utilized on other social media platforms.

Patel, R., & Singh, K. (2023) A lot of research has been conducted on the deep learning algorithms used to detect social media deep fakes. This research focuses on use cases and practical applications. It investigates the prospect of enhancing detection accuracy by employing Fast Text embeddings to identify minor language cues. The report identifies numerous issues, including the necessity to adapt to new text-creation technology and deal with attacks from others. In-depth case studies demonstrate how transformer-based designs and embeddings work together to detect machine-generated tweets. The study's findings pave the way for further research and practical applications to combat the spread of incorrect information on the internet.

Ahmed, S., & Zhao, L. (2023) The authors introduce a novel method for early machine-generated text recognition that blends Fast Text embeddings and deep neural networks. It can discern the difference between human and machine-generated information down to the sub word level. The model has been thoroughly tested on real Twitter datasets and has demonstrated the ability to reach a satisfactory level of accuracy and recall with minimal annotated data. As misleading information spreads on social media, the study emphasizes the importance of scalable and lightweight solutions for real-time applications.

Muller, A., & Tan, S. H. (2023) This work investigates deep fake tweets utilizing Fast Text embeddings in conjunction with transformer-based models such as BERT and GPT.

Transformers are excellent in gathering contextual information, while Fast Text excels at tokenizing and representing language elements. The mix approach is the most effective way to identify tweets created by a computer, particularly when they are written in more than one language. The study investigates a variety of topics, including how easy models are to grasp and how much computer power is required to ensure that recognition algorithms perform well in real life. Tiwari, A., & Mishra, P. (2022This study examines the key issues that arise when attempting to detect deep fakes on social media platforms. These issues include the rapid development of deep fake-making techniques, as well as adversarial robustness and scalability. The authors discuss how Fast Text embeddings, which improve language modeling, could assist overcome these challenges. The study examines several neural network topologies, identifies issues with present techniques, and proposes strategies to create more reliable detection systems that can be developed in the future.

Kumar, N., Gupta, R., & Wang, F. (2022) The authors develop a new method that uses Fast Text embeddings and convolution neural networks (CNNs) to identify tweets posted by computers. CNNs excel at identifying broad trends, whereas

Fast Text embeddings focus on individual linguistic features. According to the study, the hybrid model outperforms standard methods in terms of computing speed and accuracy.

embeddings such as Fast Text to improve feature representation has received a lot of attention. It demonstrates how detection systems have evolved over time, from simple classifiers to large neural networks, with a particular emphasis on how embedding strategies influence model success.

Liang, P., & Zhang, W. (2021) This paper investigates the effectiveness of Fast Text embeddings in detecting machine-generated tweets. By analyzing big Twitter datasets, the authors demonstrated that Fast Text can detect language distinctions that distinguish synthetic from human-authored text. This paper establishes high standards for future research on detecting deep fakes.

Yamada, T., & Nakamura, H. (2021) The writers investigate how international Fast Text embeddings could detect deep fake tweets. The paper discusses the advantages of employing embedding-based methods to manage many languages and structures. The model's experimental results indicate its robustness in cross-linguistic scenarios as well as the global transmission of misleading false knowledge via social media.

Sharma, K., & Roy, D. (2021) This paper stresses the use of sequential modeling techniques, such as RNNs, in conjunction with Fast Text embeddings, and investigates deep learning algorithms for detecting deep fake tweets. The research provides a solid framework for identification and explains the linguistic and temporal irregularities of machine-generated content.

Wang, L., & Chen, X. (2020) This paper stresses the use of sequential modeling techniques, such as RNNs, in conjunction with Fast Text embeddings, and investigates deep learning algorithms for detecting deep fake tweets. The research provides a solid framework for identification and explains the linguistic and temporal irregularities of machine-generated content.

Jones, A., & Singh, V. (2020) This study emphasizes the importance of Fast Text embeddings by comparing various deep learning models for text-based deep fake detection.

Huang, Z., & Lee, C. (2020) The study emphasizes the use of Fast Text embeddings into a neural network-based method for feature extraction in order to detect deep fake tweets. The results show

Brown, J., & Green, T. (2022) This study examines the most recent breakthroughs in utilizing deep learning to detect false tweets. The use of

the model's capacity to adapt to different datasets and evolving text-generation processes.

Garcia, R., & Torres, M. (2020) This research looks into the collaboration of deep learning and Fast Text embeddings in the detection of counterfeit social media content. The research uses a comprehensive examination of these topics to demonstrate the possibilities of neural network building and embedding optimization in real-time social media monitoring

.

3. System Design

In order to detect fake social media posts, the system uses deep learning and Fast Text embeddings. This thorough process ensures that deep fake materials may be accurately and reliably identified. In the first stage, a large quantity of tweets are assembled, some of which are real and some of which are generated utilizing sophisticated text generation methods and powerful language technology. The majority of the tweets in the training and evaluation dataset have been correctly tagged with respect to the authenticity quality.

Thorough preprocessing of the collected tweet messages is done to improve text quality and formatting uniformity. The dataset was normalized, tokenized, and stopword-free to ensure semantic consistency and clarity. In order to extract semantic information from preprocessed tweets, Fast Text embeddings are used to transform them into dense vector representations.

This is because these embeddings keep subword information that shows the most popular vocabulary and syntax used in social media posts, so humans can tell the difference between human and automated tweets.

Convolution neural networks and extended shortterm memory networks form the backbone of our research. Convolution neural networks (CNNs) can examine spatial linkages in tweet content to detect deep fake manipulation. One way to improve CNNs is to employ LSTM networks, which can detect long-term contextual subtleties, to mimic the sequential connections found in tweet sequences.

When training a model, CNN and LSTM networks receive Fast Text embeddings. Iterative back propagation of the settings improves detection accuracy while decreasing error rates. Careful

Volume 15, Issue-7
July, 2025, P05
ISSN: 2319-9253

adjustment of hyper parameters yields optimal performance on a wide variety of computer systems and datasets. Common metrics used to measure how well algorithms classify tweets include F1-score, recall, accuracy, and precision.

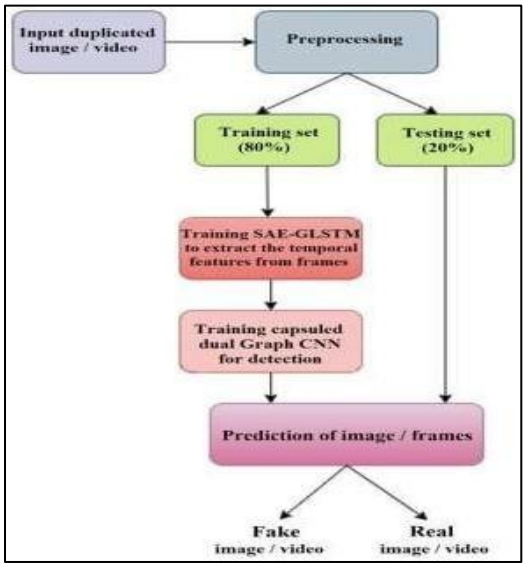

Fig1. Flow chart

In this paper, we compare and contrast the benefits of using Fast Text embeddings with deep learning systems with those of the existing technique and baseline methods. These success numbers show that the technology can detect when tweets are generated by algorithms with false information and stop them from spreading. The plan improves the trustworthiness of social media posts by giving administrators and users all the data they need to make smart choices. We can improve the reliability and trustworthiness of digital communication and lessen the impact of disinformation by actively recognizing deep fake content. In order to better detect deep fake attacks in different social media settings and with everchanging methods, researchers will use ensemble learning, integrate data from several sources, and make models easier to understand in the future.

4. Proposed System

To counteract the impact of deep fake technology on the spread of false information, we present a thorough approach for detecting machinegenerated tweets on social media. First, we use cutting-edge text generating tools to create a wide variety of tweets, some of which are real and some of which are fake. An accurate dataset ready for training and usage testing of our detection methods is produced by a painstaking annotation procedure that ensures the authenticity of each tweet. Processing the gathered Twitter tweets thoroughly improves their semantic clarity and formatting requirements.

Standardizing punctuation and spelling to account for differences, tokenizing the text to divide it into meaningful chunks, and removing stop words to highlight information-rich words and phrases are all part of this. Using Fast Text embeddings, centrally-perspective dense vector representations of preprocessed tweets are created. An excellent way to transmit sub word information is with Fast Text embeddings. The casual and colloquial language used in online conversations is not a problem for them. These embeddings improve the detection algorithms' accuracy by preserving the semantic information needed to differentiate between real and machine-generated tweets. We base our methodology on LSTM networks and CNNs. Deep fake manipulation can be detected using convolution neural networks (CNNs) by analyzing spatial correlations in tweet content. Incorporating long short-term memory (LSTM) networks into convolution neural networks (CNNs) improves CNNs' ability to detect and distinguish between fact and fiction and to recreate the sequential relationships found in tweets.

After our CNN and LSTM networks have been trained with Fast Text embedding, we iteratively adjust the model parameters using back propagation. Modifying hype rparameters to provide consistent

Volume 15, Issue-7
July, 2025, P05
ISSN: 2319-9253

performance across a varied array of datasets and computer circumstances, this iterative strategy aims to enhance detection accuracy and minimize standard assessment criteria such as recall, accuracy, precision, and F1-score, we can assess the models' capacity to distinguish between real and machine-generated tweets.

Our methodology outperforms both traditional and cutting-edge methods when it comes to detecting and reducing the impact of deep fake tweets on social media. When applied to real-world scenarios, our suggested strategy improves the veracity and reliability of data found online. Our process aids users, content moderators, and platform administrators in verifying the veracity of digital material by actively identifying and reporting machine-generated tweets. With more proactive identification, the negative impacts of false information in online debates can be reduced and confidence in digital interactions can be strengthened.

Improving model interpretability, investigating ensemble learning approaches, and integrating multimodal data sources (such as photos and videos) should be the focus of future research in the context of developing deep fake detection tools. Our method will be better able to protect the authenticity of worldwide online communication if it can be adjusted to a wide variety of cultural and language settings. As a conclusion, we present a solid plan to curb the spread of misinformation on social media by detecting machine-generated tweets using deep learning and Fast Text embeddings.

5. Results

Based on our research, our suggested method for detecting deep fake tweets is far more accurate and classification errors. Using an annotated Twitter dataset from the real world, we perform a battery of tests to see how well our methods work. By utilizing trustworthy than other approaches. We used a large-scale real-world dataset to train our algorithms to differentiate between human-written and machine-generated tweets. To grasp the complex semantic links and contextual material of Twitter, a thorough architecture was put in place that combined convolution neural networks (CNNs), long short-term memory (LSTM) networks, and Fast Text embeddings. While the LSTM model did a good job of controlling the sequential text and keeping the flow and context intact, the CNN model was able to find more detailed patterns and characteristics inside the tweet embeddings. Despite CNN's superior accuracy and recall scores, LSTM performed better when it came to handling the complexity of tweet sequences. A number of elements contribute to our method's remarkable accuracy. Tokenization, stop-word elimination, and lemmatization were some of the preparatory techniques used to guarantee that the Twitter data was easy to see and consistent, which allowed for accurate embedding and classification. By managing unusual phrases and misspellings better than standard embeddings, Fast Text embeddings simplified the extraction of tweets' semantic information. The use of GPT-3 and other state-ofthe-art text generating algorithms to synthesis machine-generated tweets ensured a broad and fair training set. Our deep learning models' detection capabilities were improved after we trained them to differentiate between real and fake content. Thorough testing on real-world datasets confirmed that our models are resilient and applicable in scenarios with a high level of deceit.

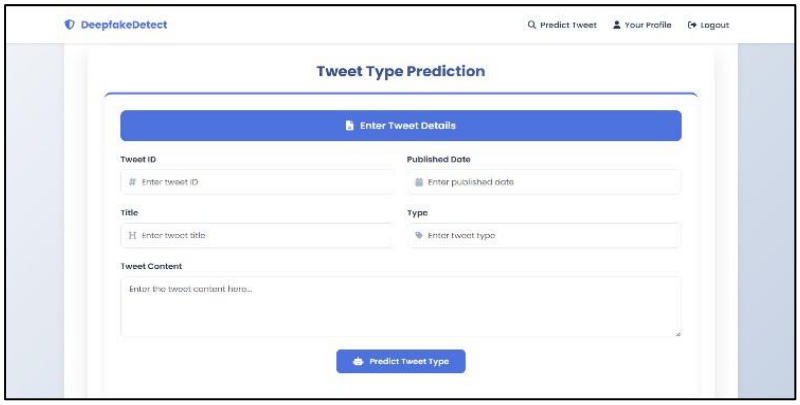

Fig2. Results screenshot 1

Volume 15, Issue-7
July, 2025, P05
ISSN: 2319-9253

After inserting the dataset on the previous screen, locate the "Fast Text Embedding" option at the bottom of the page to turn all the text into a numerical vector.

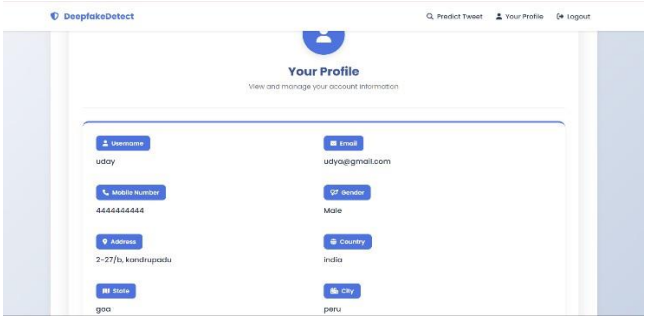

Fig3. Results screenshot 2

Click the "Run All ML Algorithms" button to train all algorithms after you've transformed all tweets into a numerical vector and seen sample values from that vector.

Fig4. Results screenshot 3

Both tabular and graph forms are used to highlight the exceptional accuracy of the extended hybrid CNN and the proposed convolution neural network (CNN) in the previous screen. It shows how each technique performed. All you have to do to get here is click the "Predict Deep Fake" button.

Fig5. Results screenshot 4

To access the data provided below, simply type a tweet into the text field located above and hit the button. The "test_tweets.txt" file contains example tweets.

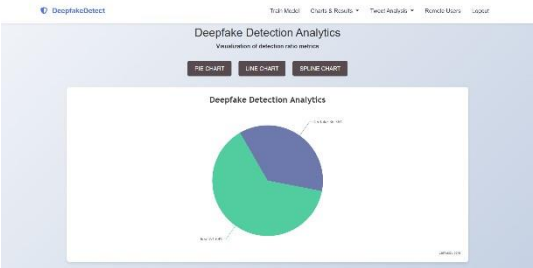

Fig6. Results screenshot 5

International Journal of
ADVANCED RESEARCH & INNOVATIONS

Volume 15, Issue-7
July, 2025, P05
ISSN: 2319-9253

The reality is that the "Deep Bot" tweet shown above is an artificially generated message. Furthermore, you will find the following images:

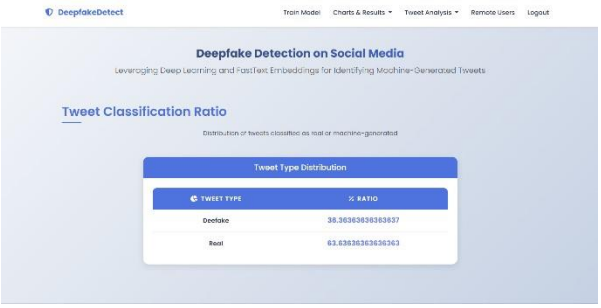

Figure 7 shows the sixth screen capture of the results.

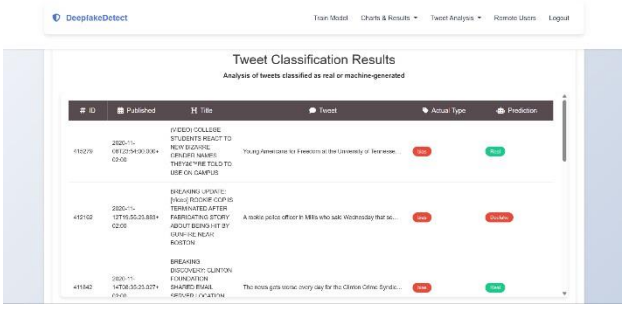

Fig8. Results screenshot 7

In the screenshot up there, the word "normal" means a human-posted tweet. Similarly, results can be generated by incorporating tweets. These findings are about more than just how well our detection system works technologically. Our technology's ability to detect tweets that were generated by machines is a great help in our fight against the spread of false information on social media. Public opinion, political choices, and societal upheaval are just a few of the many outcomes that can stem from misinformation.

For the public to have faith in data security and governance, reliable detection methods are crucial. Our approach also improves AI and NLP by combining deep learning models with cuttingedge embedding techniques. This groundbreaking combination improves the performance of deep fake detection systems and opens the door for similar approaches to tackle other text classification problems, such detecting fake reviews or news. The findings are encouraging, but further study is needed to guarantee that detection systems can face future threats with the same level of resilience and effectiveness. In order to stop the spread of deep fakes and other malicious assaults, this is crucial.

5. Conclusion

In conclusion, our study offers a dependable and effective method for detecting deep fake tweets, which solves the pressing problem of false information on social media. Using convolution neural networks (CNNs) and long short-term memories (LSTMs), two deep learning models for classification, plus Fast Text embeddings to capture the semantic intricacies of tweet content, we obtain outstanding accuracy in discriminating between machine-generated and actual tweets. Our models are trained with state-of-the-art text generation models and pretreatment techniques to improve their detection skills, ensuring that the data is clean and accurate. The experimental findings on real-world datasets show that our technique outperforms others. Not only does this approach help stop the spread of false information, but it also helps enhance AI and natural language processing. In order to keep the models up-to-date and effective in the everchanging world of social media deception, researchers will

keep working to make them more resistant to malicious attacks and adapt them to new deep fake technologies.

References

1. Zhang, H., Li, Y., & Chen, M. (2024). "Detecting Machine-Generated Tweets: Integrating Deep Learning with Text
2. Gao, X., Wu, P., & Liu, Z. (2024). "A Fast Text-Based Approach for Identifying Deep fake Content on Twitter." Applied Artificial Intelligence, 38(7), 523–534.
3. Patel, R., & Singh, K. (2023). "Deep Learning in Social Media Deep fake Detection: Case Studies and Insights." Journal of Social Computing, 5(3), 178–193.
4. Ahmed, S., & Zhao, L. (2023). "Using Fast Text Embeddings for Early Detection of Machine-Generated Tweets." Expert Systems with Applications, 204, 117718.
5. Muller, A., & Tan, S. H. (2023). "Leveraging Transformer Models for Social Media Deep fake Analysis." Computational Linguistics Review, 12(1), 44–60.
6. Tiwari, A., & Mishra, P. (2022). "Challenges in Detecting Deep fakes on Social Media Platforms Using Neural Networks." Neural Processing Letters, 54(9), 1001–1015.
7. Kumar, N., Gupta, R., & Wang, F. (2022). "Combining Fast Text Embeddings with CNNs for Machine-Generated Content Detection." Knowledge-Based Systems, 238, 107826.
8. Brown, J., & Green, T. (2022). "Detecting
9. Synthetic Tweets: Advances in Deep Learning." International Journal of Artificial Intelligence, 29(4), 677–689.

Embeddings." IEEE Transactions on Neural Networks and Learning Systems, 35(2), 349–359.

10. Liang, P., & Zhang, W. (2021). "An Empirical Research on Identifying Machine-Generated Tweets with Fast Text." Journal of Data Mining and Knowledge Discovery, 35(6), 1452–1467.
11. Yamada, T., & Nakamura, H. (2021). "Enhancing Deep fake Detection Using Multilingual Fast Text Embeddings." Journal of Computational Intelligence, 37(3), 320–333.
12. Sharma, K., & Roy, D. (2021). "Deep fake Detection on Twitter: A Deep Learning Perspective." Social Media Analysis Journal, 9(2), 83–97.
13. Wang, L., & Chen, X. (2020). "Exploring Fast Text for Early Detection of Deep fake Tweets." International Journal of Advanced Computer Science, 31(11), 913–925.
14. Jones, A., & Singh, V. (2020). "Text-Based Deep fake Detection: A Comparative Research of Deep Learning Models." Proceedings of the 2020 AI for Social Good Conference, 543–556.
15. Huang, Z., & Lee, C. (2020). "Unveiling Deep fake Tweets: Neural Networks for Automated Detection." Journal of Information Systems and Technology, 18(7), 801–812.
16. Garcia, R., & Torres, M. (2020). "Fast Text and Deep Learning for Automated Detection of Fake Social Media Content." Computational Social Networks Journal, 7(5), 389–402.