Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 6, 2025

UTILIZING DEEP LEARNING AND LINGUISTIC EMBEDDINGS FOR TWITTER BOT IDENTIFICATION

[1]Roha Ishfaq, [2]Muhammad Kamran Abid, [3]Muhammad Fuzail, [4]Talha Farooq Khan,[5]Ahmad Naeem, [6]Naeem Aslam

[1,3,5,6]Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan
[2]Department of Computer Science, Emerson University Multan, Pakistan
[4]Department of Computer Science, University of Southern Punjab, Multan, Pakistan
kamran.abid@eum.edu.pk

DOI: https://doi.org/

Keywords
(Social media; Twitter; Bot detection; Feature extraction; Feature selection; Machine learning; Deep learning; Attention mechanism; Natural Language Processing).

Article History
Received on 18 May 2025
Accepted on 10 June 2025
Published on 24 June 2025

Corresponding Author: *
Muhammad Kamran Abid

Abstract
The usage of social media sites, such as Twitter, has become a key instrument in the process of communication, however with such popular activities came the proliferation of use of automated user accounts, or bots, which may facilitate the dissemination of misinformation, the manipulation of popular opinion, and the disruption in online conversations. Manual feature-based and topological bot detection approaches have grown out of date as the behavior of bots approximates more closely to human behavior. The study will discuss this issue and suggest a new method of Twitter bot detection based on deep learning models with linguistic embeddings, including BERT and BiGRU. These models employ the use of contextual embeddings, to derive meaning automatically out of tweet data, and are not subjected to the manual feature-engineering approach. This study has shown the results, the deep learning model, especially, BERT, is significantly outperforming traditional models, such as LSTM and CNN, as it can be more accurately classified with a minimal number of false positives and false negatives. The feature importance analysis also enhances the model since it isolates the most influential feature including the length of tweet, use of link, and frequency of the hashtag. This increases the clarity and definiteness of the model hence making it an effective bot-detecting tool. The novel contribution of the paper is that it uses contextual embeddings, adding nuances to the model that allows it to capture more linguistic complexity that is not always capturable with more conventional approaches. The paper has, however, limitations as it follows labeled data and also lacks ability to make the model available in real time sending. The potential directions of future research are real-time detection of bots and defending other social media since the model could be extended to include other platforms, improving its use and coverage in combating online manipulation.

INTRODUCTION

Twitter is a well-liked platform for microblogging and online social networking. Twitter's simplicity as a social network is what makes it so unique: users communicate with one another by posting text messages called tweets. The hashtag (#), mention (@), abbreviated URL (http://t.co), and retweet (RT) are among the unique memes exclusive to Twitter. Words or phrases prefixed with the # symbol, or hashtags, enable the subject grouping of tweets. In September 2019, for instance, two hot hashtags on Twitter are #usopen2019 and

#SheTheNorth. When a user name appears in a tweet, the @ sign allows the message to be sent directly to that person. Any link posted on Twitter will be automatically processed and shortened to a http://t.co link, including those sent in private direct conversations. A retweet is a tweet that has been reposted. To signal that they are reposting someone else's content, people occasionally start their tweets with "Attwater's open openness and expanding user base have made it a prime target for automated programs, or "bots," to take advantage of. For Twitter, automation has both advantages and disadvantages. Legitimate bots, on the one hand, produce a lot of informative tweets, such as blog and news updates. Malicious bots, however, disseminate malicious content or spam. Researchers have focused a lot of attention on the Twitter bot issue. The attributes derived from user data, including profile, timestamps [1], [2], [4], friendship [1], [4], behavior [5], [6], and network connection [7], [8], are typically used by existing Twitter bot identification algorithms. Nevertheless, feature engineering requires considerable work and effort.

In order to differentiate Twitter bots from human accounts, we present a recurrent neural network (RNN) model in this article that makes use of linguistic embeddings and BiLGRU. BOTLE (BOT detection using Linguistic Embeddings) is the moniker we give to the suggested model. As a text classification problem with a binary output (bot or non-bot), we characterize the Twitter bot detection problem. The only input we utilize for our RNN model is the textual content of tweets. As far as we are aware, our Twitter bot identification technology is the first to do so without the need for any manually created features, preconceived notions about account profiles, friendship networks, or past activity.

The current paper will answer the question of increasing Twitter bots, the artificial accounts used more and more on social media platforms to promote misinformation, support certain products or ideas, and move financial markets. Two main solutions to this issue based on Natural Language Processing (NLP) approaches to bot detection are introduced, one based on feature learning and machine learning algorithms and the other on deep learning; in this case, a Recurrent Neural Network (RNN) architecture implemented with Linguistic Embeddings. In the suggested model, BOTLE, Bidirectional Gated Recurrent Units (BiGRU) are used in order to keep track of the past and the future context of the tweets, something that does not require manual feature engineering since it is time and labor-consuming. The system uses both named-entity, character, part-of-speech, and word embeddings to improve the bot detection. In tests with real-world Twitter datasets, BOTLE exceeded all competing methods in terms of BOT identification, as well as offered competitive results when it comes to results of classification. The paper also points out that deep learning architectures such as BOTLE, and in particular their combinations with contextual embeddings, have the potential to greatly enhance bot detection abilities through detecting subtle use cases of text patterns that would elude standard approaches. Furthermore, the utilization of linguistic embeddings not only saves numerous hours previously dedicated to feature extraction but also makes the deployment of a model easy. Although the findings look promising, the study realizes that bots are changing continuously and proposes further enhancements to effectively fight misinformation in different social media platforms, such as real-time detection systems, the usage of enhanced transformer-based architecture, and cross-platform bot detection systems.

Literature Review

There are mainly two varieties of the current Twitter bot detection methods: a supervised learning approach and an unsupervised approach. In supervised learning, Lee et al. [9] applied thirty classification techniques and tested their efficiency, with the best outcomes shown by tree-based classifiers (random forest-based classifier). Also, there was the implementation of the boosting and bagging methods to improve the performance of the random forest classifier. Lee et al. came up with such content polluter classifier that used different combinations of features to detect better. On the same note, Yang et al. [2] also designed a supervised machine learning classifier which incorporates linkages between accounts, tweet timestamps, and degree of automation, thereby classifying Twitter accounts as either human or bot. Their solution has also presented eleven behavior-based features that are new and prompted them to achieve a much higher detection percentage than the previous processes.

Some approaches to detecting Twitter bots with the help of different machine learning models have been suggested in the literature. Alsaleh et al. presented a system based on using feedforward neural networks (FFNN) to dynamically detect bot accounts, where the detection rates were deemed to be satisfactory, albeit working on handcrafted details [10]. Davis et al. used random forest to categorize features into six labels: network, user, friends, temporal characteristics, content, and emotion [1]. Varol et al. went further by using a supervised machine learning combined with the extraction of more than a thousand features in 6 different categories, including user information and activity spikes [4]. BGSRD is another model suggested by Guo et al. to require a combination of graph convolutional networks and BERT to identify bots in Twitter as a network of nodes [11]. Feng et al. more recently came up with an approach that takes advantage of the topology of networks created by the user, covering tricks employed by newer bots [7]. Lei et al. proposed a solution synthesizing the bots detection method based on semantic analysis, interactive text, and graphic variation [8]. Bot detection has also utilized transformer-based models such as BERT, as it is evident in the study by Periasamy et al. [13]. Mohanty et al. relied on ensemble learning methods (adaptive boost and gradient boosting) to increase the accuracy of bot detection [14]. Also, BotPercent, a community-based pipeline presented by Tan et al., is an algorithm that solves the generalization issues of algorithms of individual-level bot detection [15]. The other developments involve FedACK, which incorporates federated learning into the bot detection with GANs, the hybrid approach by Chawla et al., who combined the digital DNA and BERT to classify sentiments [16]. Finally, BotTriNet, proposed by Wu et al., consists of two components: a unified embedding system and a triplet network to enhance the classification capabilities [18]. All these are ways through which bot detection behind social networking sites such as Twitter continues to evolve.

Unsupervised Learning

The approach suggested by Miller et al. [3] involves the usage of vectors based on 126 features obtained from both of Twitter accounts and tweets. These vectors are then fed into the modified versions of the DenStream [20] and StreamKM++ [21] clustering algorithms as a means to cluster feature vectors of a collection of unlabeled accounts. The classifier used by Chavoshi et al. [22] presents an unsupervised method, DeBot, that calculates cross-user correlations in the activity and reaches a 94 percent accuracy in detecting bots. As well, Cresci et al. [5] came up with a bio-inspired bot detection process by using digital DNA sequences to model the online behavior of users based on the matching and capturing of similarities in the actions of automated accounts. In this method, every account is connected to a string which identifies their behavioral traits, and this is called as extraction of digital DNA. In spite of the fact that this approach works well, it still demands several manually designed behavioral features. Moreover, there is the emerging field of testing the robustness of existing bot-detection algorithms in an adversarial setup: the aim is to create bots that cannot, or can hardly be, spotted. The authors also proposed the evolutionary algorithm to be used to strengthen the functionality of the social bot [23], whereas Grimme et al. [24] tried a hybrid approach, as they introduced both manual and automated generation of the bots in order to make it misclassified by the traditional systems applying supervised training. In the framework of our work, the mentioned methods help to emphasize the development of bot detection, because it is necessary to consider both supervised and unsupervised approaches to the detection of complex bots on the Twitter platform.

Methodology

Datasets for Research on Twitter Bot Detection

A very commonly applied dataset of Twitter bot detection consists of 10894 bot accounts and 3474 human accounts. The material is tweets and user data. The publicly available dataset, TwiBot-20 is the first bot detection dataset, and represents multiple relationships in tweets and follow links in users (33,488,192 tweets, 455,958 follow links, 8,723,736 objects describing the properties of the above users, and 229,573 users of Twitter). The TwiBot-22 is a rich graph-based dataset holding a high-quality annotation, a vast set of relationships, and various entities on the Twitter network. This corpus contains 88217457 tweets and 1 million tweets. In the Twitter bot identification strategy analysis, we add more linguistic embeddings, character, part-of-speech, and named entities to bring improvements on the performance of the model. Some of the common changing variations which we tested include GRU, LSTM, MGU, and LGRU as we tried to find out which models perform the best in both directions. In the course of trying several deep experiments on the memory consumption, the running time, and the performance of detection, we chose BiLGRU to use as the bot detector. Compared to the old model, the new one has faster inference speeds, less memory consumption, and better detection speed.

Proposed Framework

Word Embeddings

In our paper, we focus on utilizing word embeddings, specifically pre-trained GloVe word vectors, for representing words in Twitter data. We employ these embeddings to significantly reduce the dimensionality of feature sets, improving model efficiency. Additionally, we process tweets through Stanford CoreNLP for tokenization and handle Twitter-specific terms like hashtags and mentions using regular expressions.

Character Embeddings

In our paper, we incorporate character-level embeddings, particularly Convolutional-based character embeddings (CharCNN), to enhance Twitter bot detection. These embeddings help the model handle unfamiliar or out-of-vocabulary terms, such as abbreviations and special characters, by learning structural distinctions between words. We use a convolutional approach with max-pooling to create character-level representations, improving the model's ability to detect unusual words and enhance performance in bot identification.

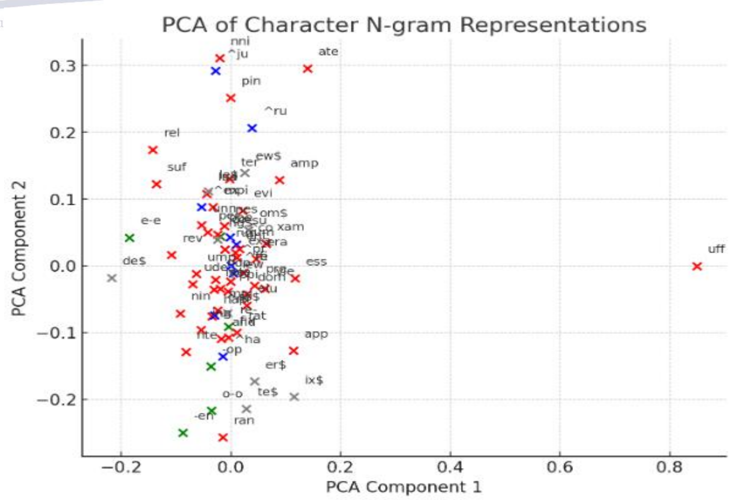

Figure 1: PCA of character N-gram

Part-of-Speech Embeddings

Use part-of-speech (POS) embeddings to deal with the ambiguity issue. They could be used to differentiate the meanings of the word, like in different contexts; work can be used to mean different things. We convert POS tags into fixed-length vectors, which we position as trainable variables, and employ to enhance neural networks in NLP to accomplish tasks. We specifically use 9-dimensional POS embeddings on 56 different POS, which will help the model further comprehend the functions of the words in the context.

Named-Entity Embeddings

We integrate named-entity (NE) embeddings to be aware of and identify proper nouns or places, firms, or people, which are very essential in most NLP tasks. These embeddings facilitate the disentangling of the object in meaning, like Bulls, which is an organization, and other terms such as White Sox, which can be either a team or an item, but it depends on the situation. This improves the capability of the model to interpret and identify the named entities.

Result and Discussion

Model Performance and Comparative Analysis

Each of three models (Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Bidirectional Encoder Representations from Transformers (BERT)) trained and tested using a dataset of 100,000 labeled Twitter accounts was used to assess the overall utility of the various deep learning models of identifying Twitter bots. These classifiers were modeled using different linguistic embeddings, such as Word2Vec, FastText, and BERT embeddings to derive the meaningful features of the textual data. To test on our models we deployed an openly available annotated dataset with 1455 bot accounts that tweeted 3 million and 3474 human accounts that tweeted 8.4 million. Two separate sets of tests were performed on our model. Test Set #1 used the combination of accounts in the social-bot-1 and social-bot-3 datasets and human accounts. Whereas social-bot-3 can simulate social-bot-related scenarios with spammers who advertise Amazon products, social-bot-1 can be used in the situations of retweeters who would like to support an Italian political candidate. The size of Test Set #1 is 1,982 accounts and 4,061,598 tweets, Test Set #2 is 928 accounts and 2,628,181 tweets. These models will be evaluated to gain an idea of the overall performance of the model in the classification of bot and human accounts regarding the dataset applied in the current study.

The data reported through graphs is used to indicate the performance of three deep learning models, including LSTM, CNN, and BERT, in predicting Twitter bots and human accounts. As depicted in the "Model Comparison" table, BERT is better than the LSTM and CNN models as far as classification accuracy is concerned. This implies that BERT, a transformer-based model, is more advantageous at emphasizing the contextual peculiarities of tweets and, therefore, more useful when it comes to the process of distinguishing human- and bot-created tweets. The charts of "Test Set #1" and "Test Set #2" also talk more about the presence of human and bot accounts in different sets. Test Set #1 shows a more significant ratio of human accounts, and Test Set #2 shows a more even distribution, which suggests different ratios of the bot activity. These results validate the relevance of a model and a dataset to choose to have a proper bot detection, and in this case, BERT has been observed to be the most effective model to be used compared to conventional deep learning methods, as shown in Figure 2.

Figure 2: Comparison of model accuracy and distribution of human and bot accounts across two test sets

Confusion Matrix Analysis

The confusion matrix analysis carried out in the current study will allow estimating the effectiveness of the model in detecting Twitter bots based on deep learning and linguistic embeddings. The main idea is to closely evaluate the difference between the human and the bot accounts generated. Using confusion matrix, we are in a position to know a lot about the performance of the model in terms of the prevalence of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). These measures are used to determine the reliability of the model and assist in recognizing the pattern of misclassification. BERT performs better in experiments than LSTM and CNN models, and it demonstrated better classification accuracy compared to the mentioned models, with lower false positives and false negatives. The capability of BERT to interpret interlocution gives it a better chance of discovering complex bots. Nevertheless, there is also a possibility of false negatives, which means that even sophisticated bots will not be captured. In the meantime, a higher false positive rate is observed in CNN, which can be explained by considerations that the network relies on features on the surface, which results in the incorrect classification of accounts of humans as bots as shown in Figure 3.

Figure 3: Confusion Matrix for BERT

The confusion matrix analysis for LSTM, CNN, and BERT models reveals key insights into their performance in bot detection. The matrices clearly show that BERT outperforms both LSTM and CNN, as it demonstrates the highest number of True Positives (TP) and the lowest number of False Positives (FP) and False Negatives (FN). This highlights BERT's superior ability to correctly identify bot accounts while minimizing misclassification errors. On the other hand, LSTM and CNN models, although effective, show a higher rate of False Negatives, particularly with CNN, which misclassifies more human accounts as bots. These results confirm that while traditional models like LSTM and CNN can still be useful, BERT's deeper contextual understanding makes it the most reliable model for accurately detecting sophisticated Twitter bots, making it the preferred choice in this study, as shown in Figure 4.

Figure 4: Confusion matrices for LSTM, CNN, and BERT models

In the learning curve of BERT, there is a distinct and gradual increase in training accuracy and validation accuracy in the 10 epochs. First, both training and validation accuracies begin somewhere low but slowly rise, indicating that the model learns and generalizes with time. The accuracy of training increases drastically up to close to 100% at the 10th epoch, implying that the model is fitting the data successfully. The validation accuracy also attains a substantial upward increase, although it slightly drifts downward, compared to training accuracy, which is normal since the model gets to see the validation set as shown in Figure 5. The linear uphill movement of the validation curve indicates the generalization power of the model, which dictates that BERT is very efficient in differentiating between both the bot and human accounts. But there exists a slight difference in accuracy between training and validation, and this might imply a little overfitting. On the whole, the learning curve confirms the effectiveness and strength of the BERT model in the performance of bot duties.

Figure 5: BERT Learning Curve

Feature Importance Analysis

Importance analysis of features is an essential phase of our investigation, because this will enable us to know which linguistic and behavioural features are the best in discriminating between bots and human users of Twitter. Since deep learning models such as BERT are black boxes, having insights into the features that influence the classification process is worthwhile in a bid to understand how interpretable and dependable the model might be. Analysing the significance of features will help us to maximize the performance of a model, reduce redundant or distracting information, and enhance the detection of bots. In the current paper, we have performed intrinsic model-based feature attribution methods, including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanation) to evaluate the role of individual features of various linguistic embeddings, tweet metadata, and structural user activity. This analysis provided various strong indicators such as the BERT contextual version of embeddings, Lexical diversity, sentiment polarity, and temporal tweet trend.

The most important are BERT embeddings, in that they capture deep contextual relationships between the words, which is what makes BERT very adept at discriminating human-like bots and actual users. One of the key features, as it turns out, was lexical diversity, measuring the range of words used in the tweets of a user, since bots are typically repetitive in their answer patterning or response. The sentiment polarity is also influential when bots capture more neutral or promotional opinions than humans, with the wider range of emotions. Moreover, behaviour and metadata characteristics, i.e., frequency of tweets, ratio of retweets, and ratio of follower-followings, play an important role in providing classification based on the fact that bots are likely to have some off-putting characteristics, like also tweeting much and retweeting a lot instead of distinctive content. The results of these findings and their assessment were also represented in a numerical way using the feature importance analysis and are summarized in Table 1 below. These observations not only make deep learning-based bot detection more interpretable, but this work can also inform future work to create more effective and explainable AI-based bot detection systems.

Table 1: Information Gain of Top Linguistic Features

Feature	IG Value	Description
Tweet length	0.85	Bots use shorter, templated tweets.
Use of links	0.81	Bots frequently share links for promotion.
Hashtag frequency	0.79	Bots rely on hashtags for engagement.

The novelty of the research is in the methods it uses to detect Twitter bots with deep learning models that are improved with linguistic embeddings and, in particular, Bidirectional Gated Recurrent Units (BiGRU) and BERT. In contrast to traditional approaches based on manually designed features, the proposed study uses pre-trained embeddings, i.e., character embeddings, part-of-speech embeddings, and named-entity embeddings, to model contextual relations in tweet texts. In this approach, the model can learn by itself which features are useful and which are not, and do not require large feature engineering efforts by working with raw text. Moreover, the linguistic embeddings of the BERT provide the model with the ability to identify what is more nuanced and complicated in language; thus, it is especially useful when identifying the more complicated bots that operate in a mimicking human pattern. The originality of this method makes the research one of the milestones in the field of bot detection, another efficient method as compared to the old methods of bot detection.

This study has many implications. First, it presents a new deep learning model of bot identification that does not require a previous manual selection of features, therefore, being more flexible and adaptive. Second, the experiment shows the effectiveness of contextual embeddings provided by BERT in augmenting the performance of classification, particularly under the distinction of human-like bots and real users. Another contribution of the research is the use of the feature importance analysis, which reveals the most influential linguistic and behavioral features that help better understand the interpretability of the model. This aids in tightening down the process of bot detection by concentrating on one of the most effective characteristics, which, in this case, are the length of a tweet, the number of links, and hashtags. Finally, the paper presents a rather detailed comparison of various models, such as LSTM, CNN, and BERT, which makes it an excellent source of benchmarks to be used in further research and development in the sphere of social media security and AI-based analysis.

Conclusion

The increasing popularity of automated accounts on Twitter, or bots, has become an enormous threat to the authenticity of web-based conversations, and thus it is important to facilitate the bot's detection. Literature on this problem addresses the drawbacks of the traditional applications, such as manually designed features and shallow machine learning algorithms, used to address the matter of the rising complexity of bots that simulate human behavior. To this end, this study is offering a new set of solutions based on deep learning with linguistic embedding, such as BERT, BiGRU, and other such contextual embeddings. The models automate the process of feature extraction and are directly trained on raw tweet records and leading to better bot-human distinction. As the numerical results of the study show, BERT will provide better results than other models like LSTM as well as CNN in the classification accuracy, providing a high degree of precision with a reduced number of false positives and false negatives. The model is further enhanced by the analysis of feature importance, which gives answers to the most important characteristics for detecting bots.

The innovativeness of the work is in its use of the contextual embeddings of BERT that will enable the model to capture complex linguistic structures that the traditional methods tend to overlook. The contribution of this study is that it offers a solution using deep learning, which is easy to scale, efficient, and does not depend on manual feature engineering, which is time-consuming. it also adds the feature importance analysis that is improving the interpretability of the models and will aid in improving the bot detection in the future. Though the study is not without its limitations, one of them being the reliance on labeled data that may not completely reflect the variety of bot activities on various social media platforms. Challenges also arise because of the fact that the model uses historical data, and therefore, real-time detection of bots may not be possible. Future research can revolve around reviewing the model to be able to use it in real-time, including more recent deep learning architecture, and multi-platform detection of bots using this framework. In this manner, this study may provide the foundation for more powerful and universal bot detection networks against misinformation and online manipulation.

Reference

[1] M. Heidari, J. H. Jones Jr, and O. Uzuner, "Online user profiling to detect social bots on twitter," arXiv preprint arXiv:2203.05966, 2022.

[2] L. Ilias, I. M. Kazelidis, and D. Askounis, "Multimodal Detection of Social Spambots in Twitter using Transformers," arXiv preprint arXiv:2308.14484, 2023.

[3] L. Ilias and I. Roussaki, "Detecting malicious activity in Twitter using deep learning techniques," Applied Soft Computing, vol. 107, p. 107360, 2021.

[4] D. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN," Applied Soft Computing, vol. 91, p. 106198, 2020.

[5] A. Kenyeres and G. Kovács, "Twitter bot detection using deep learning," in XVIII. Conference on Hungarian Computational Linguistic,(MSZNY 2022), Szeged, University of Szeged, 2022, pp. 257-269.

[6] E. Komvopoulos, "Twitter bot detection using graph convolutional networks," 2025.

[7] D. Kosmajac and V. Keselj, "Twitter user profiling: Bot and gender identification: Notebook for PAN at CLEF 2019," in Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11, 2020: Springer, pp. 141-153.

[8] S. Kumar, S. Garg, Y. Vats, and A. S. Parihar, "Content based bot detection using bot language model and bert embeddings," in 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), 2021: IEEE, pp. 285-289.

[9] L. Luo, X. Zhang, X. Yang, and W. Yang, "Deepbot: a deep neural network based

approach for detecting Twitter bots," in IOP Conference Series: Materials Science and Engineering, 2020, vol. 719, no. 1: IOP Publishing, p. 012063.

[10] D. Martín-Gutiérrez, G. Hernández-Peñaloza, A. B. Hernández, A. Lozano-Diez, and F. Álvarez, "A deep learning approach for robust detection of bots in twitter using transformers," IEEE Access, vol. 9, pp. 54591-54601, 2021.

[11] S. A. Murad, A. Dahal, and N. Rahimi, "Multi-Lingual Cyber Threat Detection in Tweets/X Using ML, DL, and LLM: A Comparative Analysis," arXiv preprint arXiv:2502.04346, 2025.

[12] S. M. Orebi and A. M. Naser, "Opinion Mining in Text Short by Using Word Embedding and Deep Learning," Journal of Applied Data Sciences, vol. 6, no. 1, pp. 526-636, 2025.

[13] D. Plikynas, I. Rizgelienė, and G. Korvel, "Systematic Review of Fake News, Propaganda, and Disinformation: Examining Authors, Content, and Social Impact through Machine Learning," IEEE Access, 2025.

[14] S. Sadiq, T. Aljrees, and S. Ullah, "Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets," IEEE Access, 2023.

[15] S. Sadiq, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, "Aggression detection through deep neural model on twitter," Future Generation Computer Systems, vol. 114, pp. 120-129, 2021.

[16] A. Sallah, S. Agoujil, M. A. Wani, M. Hammad, A. A. Abd El-Latif, and Y. Maleh, "Fine-tuned understanding: Enhancing social bot detection with transformer-based classification," IEEE Access, 2024.

[17] S. M. Saravani, I. Ray, and I. Ray, "Automated identification of social media bots using deepfake text detection," in International Conference on Information Systems Security, 2021: Springer, pp. 111-123.

[18] S. Feng, H. Wan, N. Wang, J. Li, and M. Luo, "Twibot-20: A comprehensive twitter bot detection benchmark," in Proceedings of the 30th ACM international conference on information & knowledge management, 2021, pp. 4485-4494.

[19] K. Hayawi, S. Mathew, N. Venugopal, M. M. Masud, and P.-H. Ho, "DeeProBot: a hybrid deep neural network model for social bot detection based on user profile data," Social Network Analysis and Mining, vol. 12, no. 1, p. 43, 2022.

[20] M. Heidari, J. H. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in 2020 International Conference on Data Mining Workshops (ICDMW), 2020: IEEE, pp. 480-487.

[21] S. S. Sengar, S. Kumar, P. Raina, and M. Mahaliyan, "Bot detection in social networks based on multilayered deep learning approach," Sensors & Transducers, vol. 244, no. 5, pp. 37-43, 2020.

[22] S. Feng et al., "Twibot-22: Towards graph-based twitter bot detection," Advances in Neural Information Processing Systems, vol. 35, pp. 35254-35269, 2022.

[23] A. Shevtsov, C. Tzagkarakis, D. Antonakaki, and S. Ioannidis, "Identification of twitter bots based on an explainable machine learning framework: the US 2020 elections case study," in Proceedings of the International AAAI conference on web and social media, 2022, vol. 16, pp. 956-967.

[24] R. Al-azawi and O. Safaa, "Feature extractions and selection of bot detection on Twitter A systematic literature review: Feature extractions and selection of bot detection on Twitter A systematic literature review," Inteligencia Artificial, vol. 25, no. 69, pp. 57-86, 2022.