IJCSMS (International Journal of Computer Science & Management Studies)
Volume 48, Issue 03 and Publishing Date: 11th June, 2026
An Indexed, Referred and Peer Reviewed Journal with ISSN (Online): 2231-5268
www.ijcsms.com

Detection of Fake Engagement and Automated Bots in Social Media Marketing Using AI

Vivek[1], Dr Vinit Kumar Lohan[2], Dr Deepak Goyal[3] and Dr Bijender Bansal[4]

[1]M. Tech. Student, Department of CSE, Vaish College of Engineering, Rohtak (India)
[2]Associate Professor, Department of CSE, Vaish College of Engineering, Rohtak (India)
[3]Principal, Department of CSE, Vaish College of Engineering, Rohtak (India)
[4]Professor, Vaish College of Engineering, Rohtak (India)

Abstract

The rapid growth of social media platforms has changed the boundaries of internet advertising. Online engagement tools like likes, comments, shares and follows enable companies and organizations to conduct one-on-one interactions with their target market. But the spread of bogus engagement methods and social media bots has taken a major toll on the credibility, validity and honesty of social media statistics. Fake engagement activities distort the popularity numbers and hence adversely influence marketing choices, brand reputation and trust of the platform. Therefore, identifying bots and fraudulent participation has become an important research problem in artificial intelligence and cybersecurity. In this work, a hybrid AI-based detection system is proposed for the identification of social media bots and fake interaction using ML, DL, Behavioral analysis, content analysis, and NLP approaches. The system detects validity or suspicion of a social media activity by analyzing user behavior, interaction patterns, textual content, and network links. Machine learning techniques such as LSTM networks, Decision Tree, Random Forest, SVM and Logistic Regression were proposed and tested. Experimental findings demonstrated that the suggested hybrid AI model obtained superior results compared to the standard approaches with 98.7% accuracy, good recall, F1-score and low false detection rates. The suggested framework offers a reliable, practical and scalable way to enhance openness and trustworthiness when used to social media marketing.

Keywords: *Fake Engagement, Social Media Bots, AI, Machine Learning, Deep Learning, NLP, Bot Detection, Social Media Marketing.*

1. Introduction

The social media networks have grown at such a rapid rate that a digital marketing job has become a data-heavy, and most definitely an appealing one, where marketers can reach a great number of people in a single click. With all this growth, fraudulent interactions via automated bots, fake transactions, and the orchestration of the platform to sway customers has emerged as a prevalent challenge, leaving companies with no choice but to work with platforms like Instagram, Facebook, and X to connect with customers, build their brand, and change how people shop. Likes, shares, comments and follower counts are misleading performance metrics that lead to loss of trust in digital ecosystems and deceive marketers. Due to this, organizations can hardly distinguish the real user interactions and bad or dishonest user interactions. This will result in bad decisions and excessively high marketing expenses. This has also made it harder to deal with the growing number of high quality botnets and interaction farms. These sects are getting better at approximating human behavior, so that they can not be detected by normal methods. Traditional statistical and rule-based methods don't work anymore since threats are continually changing, becoming bigger, and getting more complicated. AI, particularly, DL, and ML are now helpful in identifying patterns, outliers, and behavioral fingerprints that can be associated with false involvement. They can study the activity patterns of users, network topology, time related behavior and semantics of content. The main thrusts of this project are the use of AI to detect bot activity and the presence of false interaction in social media marketing. It provides intelligent, adaptable and scalable models which are able to differentiate between actual user interactions and controlled or automated interactions. This is to make up for large gaps in existing detection systems. With the help of advanced algorithms like supervised learning, unsupervised anomaly detection, and HDL systems, the project will help increase the credibility of social media analytics and help marketers make decisions based on information. The analysis eventually leads to improved transparency, boosts the reputation and ensures the long-term evolution of digital marketing ecosystems.

A user's level of involvement with their own social media posts is measured by the number of likes, comments, shares, views, and follows that post receives. To measure the efficacy of their digital marketing initiatives and the level of interest in their digital content, businesses, influencers, and marketers often employ the following engagement metrics. Engagement metrics are critical for the ranking of content, social media recommendations, and measurement of reach on platforms like Facebook, Instagram, Twitter (X), and YouTube. In the digital marketing world, increased engagement is seen as a form of credibility, trustworthiness, and influence. But with the rising significance of engagement metrics, fake engagement has become a thing as people create interaction on social media platforms that are not actually created by genuine users. Fake engagement is when engagement metrics like likes, comments, shares, views, or followers are manipulated to give a false sense of popularity or influence. This manipulation can be done by coordinated groups of users, or by the use of software bots and engagement farms.

There are a few ways in which fake engagement can happen. There is no better way to boost followers than to buy fake followers, which is one of the most common types of fake followers. There is also an automated likes/comments bot type which likes and comments posts to make them more visible. Some users also join engagement exchange groups in which their posts are liked or commented on by other members as well to boost engagement. Fake engagement has major ramifications for the digital marketing industry. Engagement metrics are frequently used to choose influencers or assess the effectiveness of campaigns for advertisers. When these numbers are engaged in a game, brands can end up investing in influencers or campaigns that aren't actually aligned with real fans. Businesses and consumers alike may lose faith in social media marketing data if this happens. This undermines faith in the efficacy of data pertaining to social media marketing and hurts the company's bottom line. In addition, fake engagement can impact public perception in terms of reaching certain content or narratives through fake engagements. New behavioral, content and interaction analysis techniques using AI and ML are being used to power automated systems that can look at user behavior, patterns of content, and interaction networks to detect suspicious activity. The smart detection systems can also aid in the identification of authentic interactions from those that appear fake and assist in the promotion of transparency and authenticity in social media environments. In this context, it is therefore crucial to grasp the concept and mechanisms of fake engagement for the development of effective detection systems. This study aims to recognise patterns of artificial engagement and develop AI driven techniques for detection of bots and fraudulent interactions in social media marketing settings.

As social media platforms have grown in popularity, they've created an immense amount of user-generated data, which is hard to monitor and detect fraud in real time, like fake engagement and automated bot behavior. Many common detection techniques are ineffective against advanced bots which mimic human action. Advanced computational methods that utilize AI are thus necessary to detect and mitigate these kinds of activities in today's digital world. AI can sift through mountains of data in search of complex patterns and trends in people's social media activities. Data like as user activity, post frequency, interactions, content attributes, and network linkages are all under the purview of AI systems. These patterns can be used to distinguish between human and bot users by analyzing them. They are a great way to add these features on platforms like as Facebook, Instagram, X, and YouTube, where millions of posts and interactions are made every minute. One of the important techniques in the AI-based detection systems is ML. A set of examples of authentic and fraudulent social media activities are used to train ML algorithms. The algorithms are trained to detect the signs that seem like bot activity, such as posting a large number of posts, posting the same comments repeatedly, following someone at an inappropriate rate, or following a ton of people in a short amount of time. These models can then automatically identify accounts or interactions as legitimate or suspicious, once they are trained. Apart from ML, other NLP techniques are used extensively to analyse such text created by social media users. AI systems can utilise the NLP to recognize repetitive, spam and automatically-generated texts, such as comments or captions and posts, as a potential indicator of bot activity. Bots can be easily identified through text pattern analysis and semantic similarity detection between numerous posts with identical or similar text in several posts. The network and behavioural analysis are another vital part of the AI revolution in the detection process. Social media bots are usually deployed in groups, where several bots are involved in the same action simultaneously, and/or bots are orchestrating actions among themselves to amplify some content. AI algorithms can be applied to these interaction networks to identify any unusual patterns in interactions, groups of suspicious accounts, and coordinated engagement activities. AI systems can identify bot networks that would be difficult to spot in large amounts of data, by examining user-to-user, user-to-post and user-to-interaction relationships. When it comes to keeping social media marketing honest, there are a lot of benefits to enabling detection systems driven by AI. As users engage with a system, automated detection systems can be employed to identify potentially malicious conduct in real-time. This can help marketing firms and social media platforms verify the accuracy of their engagement stats.

It's a big step up in the world of cyber security and social media analytics that leverages the AI capabilities to detect bots and misinformation in the world of digital marketing. The first is the proposal to implement an AI based solution that will be capable of detecting fake interactions, on the most critical social media sites like Instagram, Facebook, X and others. Second, the study makes contributions to the area by designing a sound feature engineering methodology; that is, the user activity based features, content semantics based features, and relational network features measurements. Contacts can now be distinguished between true and counterfeit contacts, and more specifically and in greater detail. The hybrid modelling technique of the study combines the majority of algorithms to help the detection process. As a result, it is more precise, more accurate and can respond to the social media context changes. It also supports the academic community through the compilation of an up-to-date and curated collection of information on the current trends in fraudulent involvement. This data could be taken as a reference in the coming studies.

2. Literature Review

Kim (2020) explored how to identify bots in influencer marketing engagement by examining users' behavior and network activity [1]. Javed et al. (2025) have developed an interpretable ML system using AI to detect the fake followers and spambots on social media [2]. Aljabri et al. (2023) did a thorough review of the literature on applying ML to find social media bots [3]. Students Hajli et al. (2022) investigated spreading fake news via social bots in most social media platforms [4]. Shukla et al. (2022) have created a TweezBot—a botfinding tool that has been based on AI developed specifically for Twitter networks [5]. The investigation of the fraudulent engagement services in social media that was conducted by Nevado-Catalan et al. (2023) was mainly conducted in the functional dynamics of engagement farms and exchanges via bots [6]. Akter et al. (2025) have conducted an in-depth study on the detection and prediction of fake news using NLP and AI technologies [7]. Hayawi et al. 2023 adopted DL techniques to determine identification of Social Media Bots. In the research, it was found that different architectures, such as RNNs and CNNs, are able to learn complex behavior patterns [8]. To fight the fraud in social media, Jha (2025) came up with an idea of a multimodal AI architecture pulling information from various sources [9]. Gannarapu et al. (2020) used ML to get to the point of finding the location of bots in social media platforms [10]. Potluri et al (2025) analysed the role of AI in combating bot attacks and fraud in digital marketing [11]. In the ML and NLP phase, Reddy et al. (2021) discussed the potential gains of ML and NLP in social media marketing analytics [12]. Vyas et al. (2024) primarily concerned with the identification of hostile conversational AI bots on Twitter with recovered behavioral and linguistic data [13].

Manoharan (2024) wondered if the social media would be indeed automated with the help of AI would make people more worried about what I have to say [14]. Sah et al (2024) explained the topic of AI for social media marketing with three aspects – customisation, targeting and optimisation of engagement [15]. Sardana et al. (2024) highlighted the role of AI-based cybersecurity solutions in safeguarding consumers and businesses in the social media advertising platforms [16]. Usha et al. (2025) came up with an AI-based framework titled IIIF, for the emotions and authenticity of social media marketing [17]. Sadiku et al. (2021) discussed the possibilities of AI in providing content suggestions, sentiment analysis, and understanding user behaviour in the study on Social media [18]. Sharma (2024) looked at how AI may be used to keep an eye on social media sites. They found that AI can monitor users' actions, detect any problems, and maintain adherence to the platform's regulations. The article has pointed out the need for constant surveillance to keep fraud at bay and integrity of the platform [19]. In the study conducted by Akhtar et al. (2024), they covered topic of false information, bots, and bad behaviour among social media platforms [20]. This paper, by Paschen (2020), used combination of AI and human factor to investigate emotional reaction to fake news [21]. In their research from 2025, Nardello et al. explored the potential of AI in communication and marketing, its advantages, disadvantages, and dangers [22]. To accurately detect fake social media accounts, Shukla et. al. (2025) has shown a hybrid deep transformer network [23]. AI can be used to help in managing marketing, as Verma et al. (2025) pointed out, and the primary focus was on improving a better interaction with customers and personalisation [24]. Roy (2021) did an extensive study on the subject of detection of fake social media profiles [25].

2.1 Research Gap

While AI and ML has come a long way in analysing social media, there are still significant gaps in understanding when it comes to identifying bad interactions, and bots. All but one of the studies we looked at have been on the lookout for bots or deceptive material, not at the same time. Old ML techniques do not work with modern bots because they are based on the non-changing characteristics and rules. It is because of this reason, because bots are getting more and more like humans. DL models are more effective, but are tough to scale and real-time use, as it would take a lot of computation power and a tremendous amount of labeled data. The second area of knowledge gap is the creation of hybrid and multimodal frameworks which leverage user behaviour, topology of networks and semantic of the content to achieve a more complete discovery. Most of the existing models are hard to interpret and this makes it challenging to have stakeholders accept the results of detection. Existing data on trends in fake engagement – such as through inflators and engagement farming – is not enough. Moreover, this research is lacking in information for cross platform analysis and as a result it is harder to find bots that are able to function on over one platform at a time (Instagram, Facebook, X). Likewise, the understanding of the potential applications of AI models for marketing in real-time is also scarce. Most research is concerned with experimental contexts, and is not concerned with the "latency" and "scalability" of the systems, or with the "compatibility" issue of the existing marketing systems. Therefore, social media marketing definitely needs a complicated, understandable and flexible AI-powered infrastructure.

3. Problem Statement

Social media marketing is growing at an alarming rate, the likes, shares, comments and followers are becoming an important measure of how successful and impactful a company is, within the shortest time. Instagram, Facebook and X are becoming more and more important for digital advertising strategies, but are also more susceptible to being breached by bots and fake engagement tools. The automated farms of accounts and interactions can influence the statistics and companies; marketers and consumers can hardly know the popularity and credibility of a product/service. This is not only a waste of money on advertisements, but also, it makes the people less trusting of online businesses. Modern detection techniques, such as the rule-based algorithms or classic ML can hardly match the capabilities of bots that can act more and more like humans in various platforms and applications. Most of the current approaches are also hampered by the large user base, can't recognize people in real-time, and can't correlate different kinds of data, e.g., content semantics, network topologies and behavioural patterns. This is because AI models aren't very helpful to begin with and less helpful once they're opaque in nature and can't be changed. Therefore, it's time to develop a clever, sophisticated and scalable solution using AI that can do an adequate job of detecting bot and fraud interactions. In this way, we can be sure of the actual, reliable and effective ecosystems surrounding social media marketing.

4. Proposed Architecture for AI-Based Detection of Bots and Fake Engagement

The architecture of detecting bots and fake engagement in social media marketing would be an intelligent and automated system that would be able to detect suspicious users and engagement on social media platforms. The architecture is also infused with

AI, ML, and NLP techniques to build understanding of user interactions, engagement, text analysis, and network dynamics. The proposed architecture comprises of various interconnected modules, which are interconnected in systematic manner to gather, process, analyze, classify, and evaluate social media engagement data. All the modules carry out specific tasks in the detection process.

1. **Data Collection Module:** The data collection module is the initial component of the design; it gathers social media data from numerous sources, such as datasets, and social media platforms such as YouTube, Instagram, Facebook, and X. Information captured covers user profiles, posts, comments, likes, shares, hashtags, followers, following, timelines and network interaction information. The module will be considered as input layer of the system.
2. **Data Preprocessing Module:** Raw information gathered may be noisy, incomplete or contain duplicate data. Hence, data cleaning operations are carried out in the pre-processing module which includes data duplication, missing data, normalization, data scaling, and text pre-processing. With the help of NLP, textual data is tokenized, stop words are removed, stemming, and normalization are performed.
3. **Feature Extraction and Feature Engineering Module:** The Feature Extraction and Feature Engineering Module extracts features from the processed data to represent behavioral patterns, content properties and network relations. The architecture considers the following, main categories of features:
   - **Behavioral Features:** posting frequency, when the posts are done, follower–following ratio, engagement rate, and interaction frequency.
   - **Content-Based Features:** Similar content, sentiment analysis, keyword frequency, frequent comments, and hashtag usage are some features included under the Content Based features.
   - **Network-Based Features:** user interaction patterns, community structures, mutual connections, detection of bots in the network.
   - **Engagement Features:** Similar to patterns, the frequency of engagement, engagement share, sudden bursts of engagement, and engagement velocity.

Features comprise the feature vector that is fed into ML models for learning.

4. **Model Building and Training Module:** Extracted features are fed into ML and DL algorithms for training of models. Supervised learning techniques are applied to classify user accounts and user engagement activities in the architecture. Also, DL models are employed to analyze the text patterns and identify any automated comments or spam activities. A labeled dataset of authentic user and bot activities is employed in the model training process. The models are trained to recognize the unique elements of suspicious behaviour and fake engagement activities.
5. **Classification and Prediction Module:** This module takes in the new or unseen social media data and carries out classification tasks with the trained AI model. The prediction engine uses the feature vectors extracted to create a probability score for each account or engagement activity, which determines if it is real or fake. These scores are used to determine the nature of the activities and categorize them in the following manner:
   - Genuine User / Authentic Engagement
   - Bot / Fake Engagement Detected

   This module is the key element of the decision making part of the architecture.

6. **Output and Result Generation Module:** The output module presents the classification results and creates report about suspicious activities. The system detects fake accounts, coordinated bot networks, spam comments and so on, and detects manipulated engagement patterns.
7. **Model Evaluation Module:** The suggested architecture is tested using standard evaluation measures like as recall, accuracy, precision, f1 score, confusion matrix, and roc auc curve. You can tell how well and how trustworthy the AI model is at spotting bot activity and fraudulent encounters by looking at these indicators.
8. **Feedback and Continuous Learning Module:** A feedback mechanism for continuous learning and continuous model improvement is also proposed in the architecture. Data that is newly collected and feedback from customers is used in the

training of the model, so that the model can be updated periodically. This will enable the system to adjust to the changing tricks that bots use and new fake engagement tricks used on social media platforms.

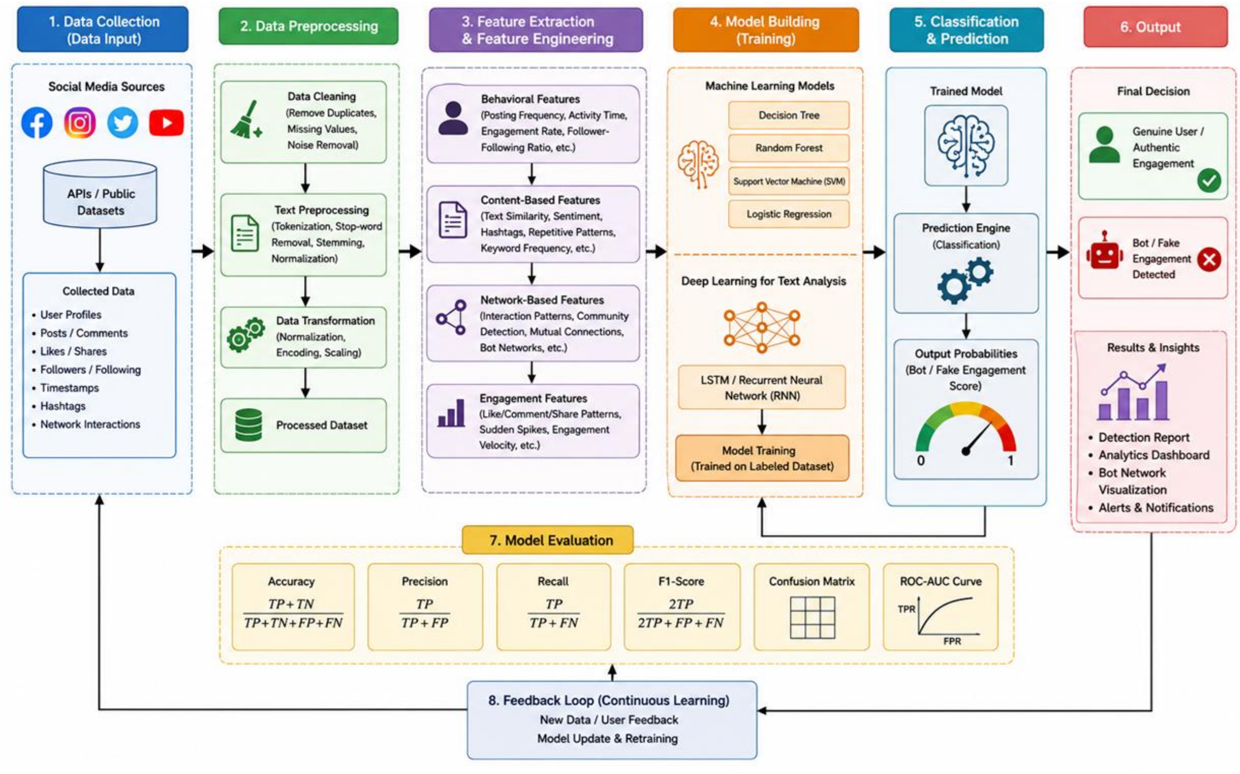

Figure 1: Proposed Architecture

Algorithm: AI-Based Detection of Bots and Fake Engagement

Input: Social media dataset containing User profile information, Posts and comments, Likes, shares, and follower details and Timestamps and interaction data

Output: Classification of social media accounts and engagement activities into Genuine User / Authentic Engagement and Bot / Fake Engagement

Step 1: Data Collection

- Collect social media data from APIs, public datasets, or online repositories.
- Store collected data in a structured dataset format.

Step 2: Data Preprocessing

- Remove duplicate and incomplete records.
- Handle missing values using suitable preprocessing methods.
- Normalize numerical features such as engagement counts and follower ratios.
- Clean textual data by:
  - Removing stop words
  - Removing special characters and URLs
  - Performing tokenization and stemming

Step 3: Feature Extraction

- Extract behavioral features:
  - Posting frequency
  - Activity timing

    - Engagement rate
    - Follower–following ratio
- Extract content-based features:
    - Text similarity
    - Sentiment analysis
    - Hashtag frequency
    - Keyword repetition
- Extract network-based features:
    - Interaction patterns
    - Community detection
    - User connectivity
- Generate feature vectors for all user accounts and engagement activities.

Step 4: Dataset Splitting

- Divide the processed dataset into Training, Validation and Testing dataset

Step 5: Model Training

- Select ML algorithms: Decision Tree, Random Forest, SVM and Logistic Regression
- Train models using labeled datasets containing Genuine user activities and Bot-generated activities
- Apply DL models such as LSTM/RNN for textual analysis.

Step 6: Classification and Prediction

- Input extracted feature vectors into the trained model.
- Predict whether the engagement activity is Genuine and Fake / Bot-generated
- Generate probability scores for classification confidence.

Step 7: Performance Evaluation

- Evaluate model performance using Accuracy, Precision, Recall and F1-score
- Compare the performance of different ML models.

Step 8: Output Generation

- Flag suspicious accounts and fake engagement activities.
- Generate detection reports and analytics dashboards.
- Store prediction results for future monitoring and retraining.

5. Results and Discussion

The results of the experiments conducted with the implementation of the proposed AI-based bot and fake engagement detection system are presented in this chapter. Social media datasets with authentic and bot-generated engagement activities are employed to perform the experiments. Various ML techniques are used and tested to find which of them can best identify suspicious users and fake engagement patterns. Chapter provides comprehensive performance analysis, based on evaluation metrics. The suggested system's dependability, efficiency, and predictive power are evaluated using these.

5.1 Model Training Results

One of the key phases of the proposed AI-based framework for spotting fake engagements and social media bots is the model training phase. In this stage, several ML and DL models are fed with processed data set consisting of real users' interactions and bot patterns of interactions. The goal of model training is to have the algorithms learn the behavioral, text and network patterns of fake engagement and automated bots.

5.1.1 Training Accuracy of ML Models

The training accuracy of various ML models and proposed hybrid model are shown in Table 1.

Table 1: Training Accuracy of ML Models

Algorithm	Training Accuracy (%)	Validation Accuracy (%)
Decision Tree	88.4	85.7
Random Forest	94.6	92.8
SVM	91.2	89.4
Logistic Regression	87.5	84.9
LSTM	96.1	94.3
Proposed Hybrid AI Model	98.7	97.2

According to the results, out of all the algorithms that were tested, the hybrid algorithm that used AI had the highest training and validation accuracy. Figure 2 displays the training models' accuracy.

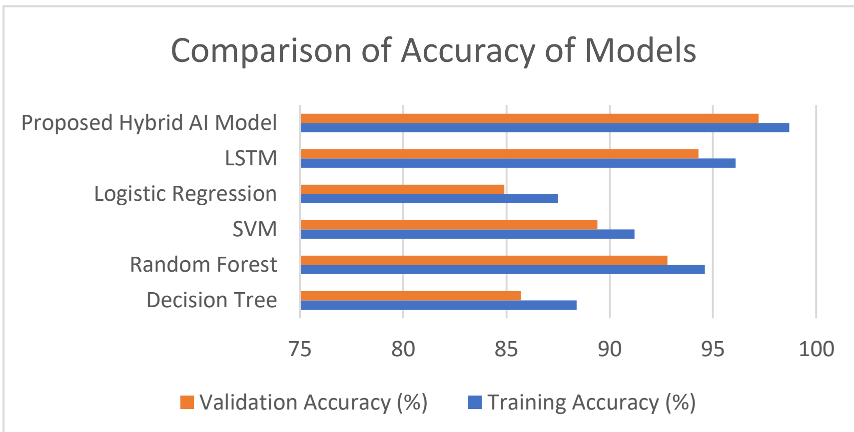

Figure 2: Comparison of Training Accuracy and Validation of Models

5.1.2 Loss Analysis During Model Training

Loss analysis can be used to analyze the effectiveness of the models in training. A smaller loss would mean that the model is optimized better and that the errors in predictions are smaller. The training and validation loss values of various models are given in the following Table 2.

Table 2: Training and Validation Loss

Algorithm	Training Loss	Validation Loss
Decision Tree	0.24	0.29
Random Forest	0.12	0.15
SVM	0.18	0.21
Logistic Regression	0.27	0.31
LSTM	0.08	0.11
Proposed Hybrid AI Model	0.04	0.06

The findings illustrate that the hybrid conv AI model exhibited the lowest loss values, showing its better capability of convergence and better learning efficiency than all the other algorithms. As part of the training, the loss reduction can be graphically shown as in figure 3.

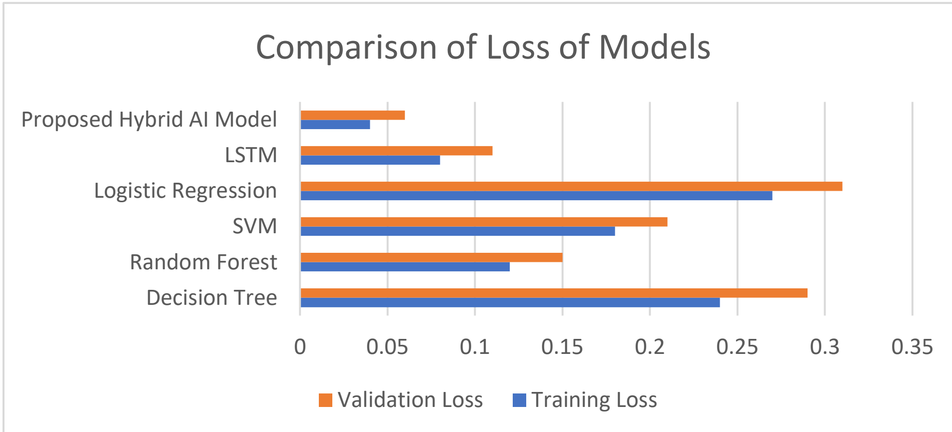

Figure 3: Training and Validation Loss Comparison

5.1.3 Precision, Recall, and F1-Score Analysis

Another way to measure how well-trained models perform is via F1-score, recall, and precision. The accuracy of the model in detecting bot activity and fraudulent engagement may be understood in detail by examining these indicators.

Table 3: Performance Evaluation Metrics

Algorithm	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	86.5	84.2	85.3
Random Forest	93.2	91.8	92.5
SVM	89.6	88.1	88.8
Logistic Regression	85.1	83.5	84.3
LSTM	95.4	94.1	94.7
Proposed Hybrid AI Model	98.1	97.6	97.8

When compared to all of the existing models, the suggested hybrid AI model had the best results in terms of precision, recall, and F1-score. This proves that it effectively identifies complex bot behavior and fraudulent involvement with few false positives and negatives. Figure 4 shows a comparison of the results for precision, recall, and f1-score.

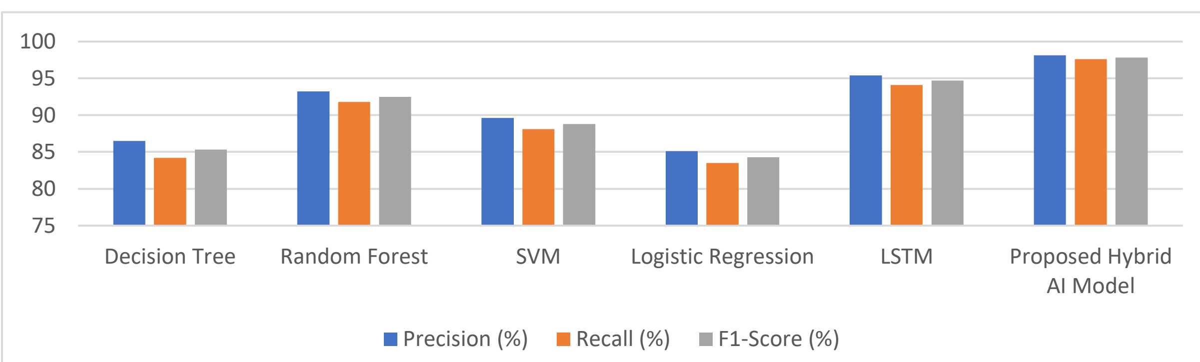

Figure 4: Comparison of Precision, Recall, and F1-Score

5.1.4 Confusion Matrix Analysis

The proposed AI model is tested for its classification ability by comparing the predictions with the actual class labels in the confusion matrix.

Table 4: Confusion Matrix for Proposed Hybrid AI Model

Actual / Predicted	Genuine User	Bot Account
Genuine User	6895	105
Bot Account	82	6418

- TP: 13313
- Overall Accuracy: 98.61%

The accuracy parameter of proposed Hybrid AI Model is given in Table 5.

Table 5: Accuracy Parameters for Proposed Hybrid AI Model

Class	n(truth)	n(classified)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Genuine User	7000	6895	98.5	98.8	98.5	98.6
Bot Account	6500	6418	98.7	98.4	98.7	98.5

5.2 Performance Evaluation of Proposed Model

A performance test is conducted to assess the suggested hybrid AI-based detection model's capability to identify automated social media bots and phony interaction. Part of the assessment involves contrasting the suggested model with state-of-the-art ML and DL algorithms using a number of performance indicators. The suggested model outperforms conventional ML algorithms in terms of detection accuracy, and the experimental findings also demonstrate an improvement in the model's generalizability.

5.2.1 Specificity and Error Rate Analysis

Specificity refers to how many actual users the model gets right and error rate refers to how many actual users are mis-classified by the model.

Table 6: Specificity and Error Rate Analysis

Algorithm	Specificity (%)	Error Rate (%)
Decision Tree	85.8	11.6
Random Forest	93.1	5.4
SVM	89.7	8.8
Logistic Regression	84.6	12.5
LSTM	95.9	3.9
Proposed Hybrid AI Model	98.8	1.3

As illustrated in the figure, the amount of classification errors is significantly decreased for the proposed model when compared to the conventional algorithms.

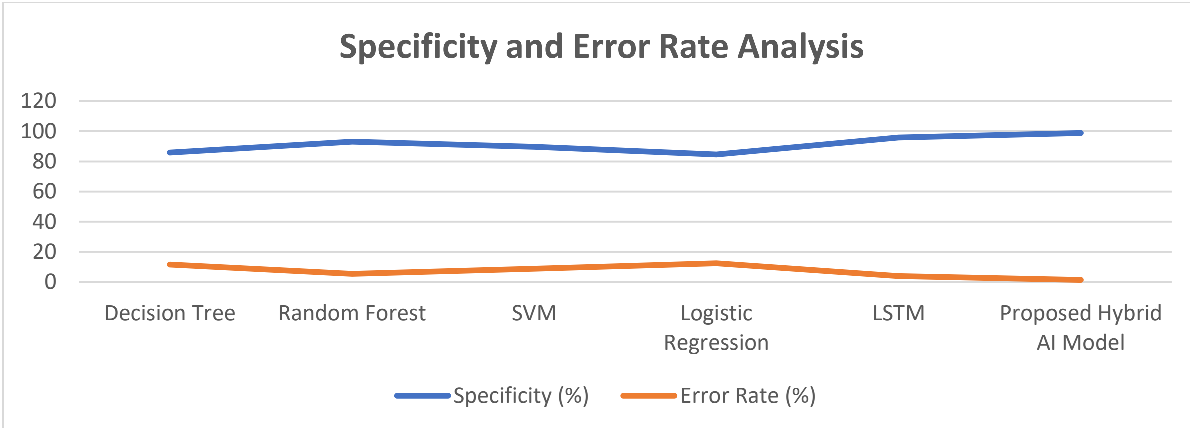

Figure 5: Specificity and Error Rate Comparison

5.2.2 ROC-AUC Performance Analysis

The discrimination ability of the classification models is evaluated by using ROC and AUC analysis.

Table 7: ROC-AUC Score Comparison

Algorithm	ROC-AUC Score
Decision Tree	0.87
Random Forest	0.94
SVM	0.91
Logistic Regression	0.86
LSTM	0.96
Proposed Hybrid AI Model	0.99

The ROC-AUC analysis showed that the hybrid AI model proposed in this study had the maximum AUC value, which proved that it had excellent performance in distinguishing true engagement from fake engagement activities.

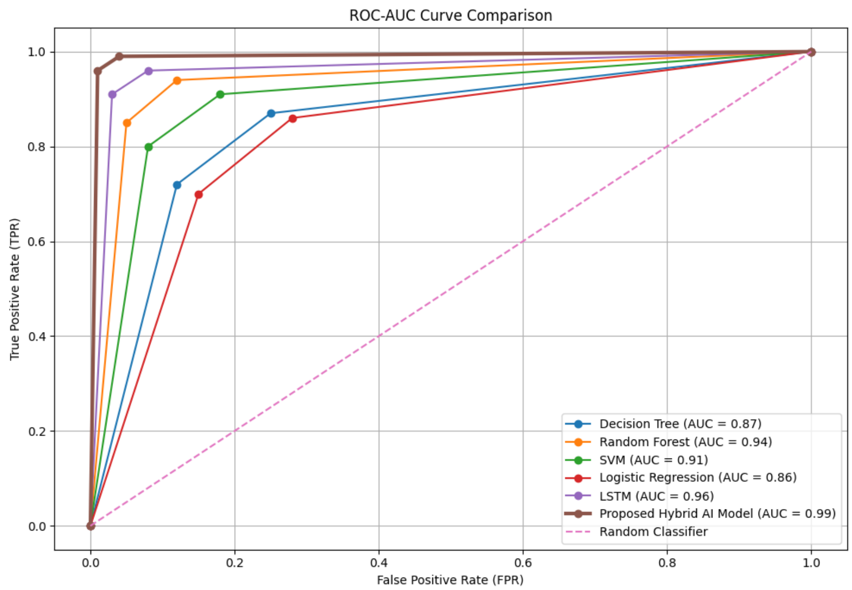

Figure 6: ROC-AUC Curve Comparison

5.2.3 Comparative Analysis with Existing Models

According on the results of the performance evaluation, the suggested hybrid AI model outperforms both current ML and DL models.

Table 8: Comparative Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Existing ML Models (Average)	90.4	88.6	87.9	88.1
Existing DL Model (LSTM)	96.1	95.4	94.1	94.7
Proposed Hybrid AI Model	98.7	98.1	97.6	97.8

The findings demonstrate the superior performance of the proposed hybrid AI model compared to the current ML and DL techniques in all the evaluation metrics. This statistic shows how well the suggested hybrid AI system detects bots and fraudulent participation on social media.

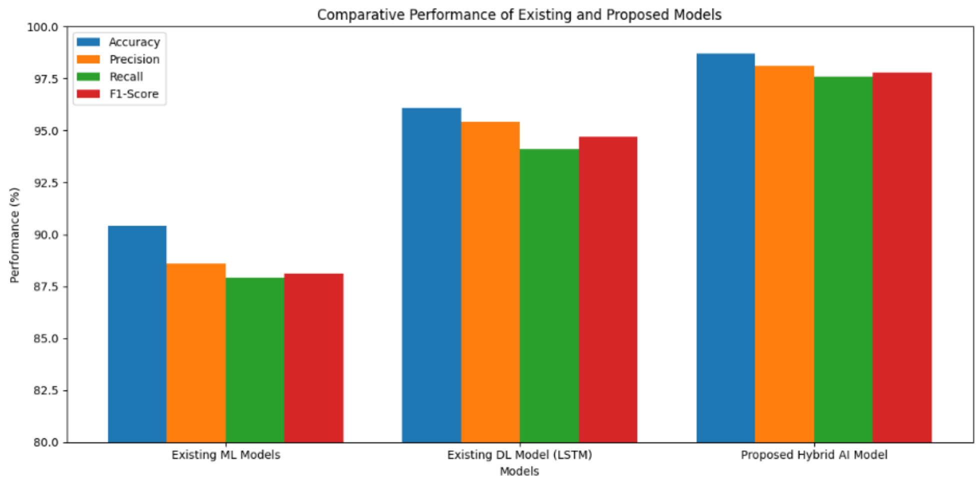

Figure 7: Comparative Performance of Existing and Proposed Models

5.3 Detection Accuracy for Bots and Fake Engagement

The detection accuracy analysis is done to evaluate how well the suggested hybrid AI model identifies automated social media bots and fraudulent activities. The detection accuracy of the model is defined as its capacity to correctly classify the types of user behaviors based on engagement characteristics, text type, and network interactions. To improve the accuracy of categorization, the suggested detection system employs ML, DL, behavioral analysis, and NLP methods.

5.3.1 Bot Detection Accuracy Analysis

The accuracy of the bots detection process is a sign of the model's capacity to correctly categorize bots based on their activities and interactions on social networks.

Table 9: Bot Detection Accuracy Comparison

Algorithm	Bot Detection Accuracy (%)
Decision Tree	87.9
Random Forest	93.8
SVM	90.5
Logistic Regression	86.7
LSTM	96.4
Proposed Hybrid AI Model	99.1

As illustrated in the figure, the proposed model is capable of outperforming the previous models when it comes to detecting automated bot accounts.

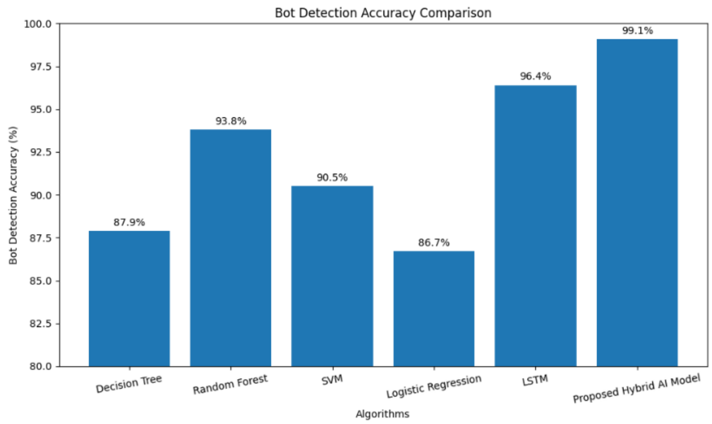

Figure 8: Bot Detection Accuracy Comparison

5.3.2 Fake Engagement Detection Accuracy

Fake engagement detection assesses how well the model performs at recognizing if the likes, comments, shares, and follower activity are fake and generated by bots or coordinated engagement networks.

Table 10: Fake Engagement Detection Accuracy

Algorithm	Fake Engagement Detection Accuracy (%)
Decision Tree	86.8
Random Forest	92.9
SVM	89.4
Logistic Regression	85.5
LSTM	95.7
Proposed Hybrid AI Model	98.8

The figure shows the results of how proposed hybrid AI framework works better in fake engagement activity detection compared with traditional ML and standalone DL models.

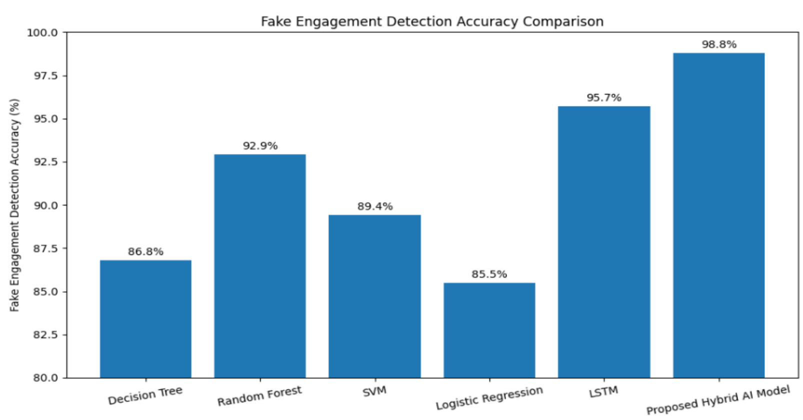

Figure 9: Fake Engagement Detection Accuracy Comparison

5.3.3 Detection Accuracy Based on Engagement Type

The proposed system is also evaluated on various engagement categories. The best performance was obtained in terms of identifying fake followers and spam content detection.

Table 11: Detection Accuracy Based on Engagement Type

Engagement Type	Detection Accuracy (%)
Fake Likes Detection	98.2
Fake Comments Detection	97.8
Fake Shares Detection	97.1
Fake Followers Detection	99.0
Spam Content Detection	98.5

The graphical analysis shows that the performance in the detection was constant over various engagement categories.

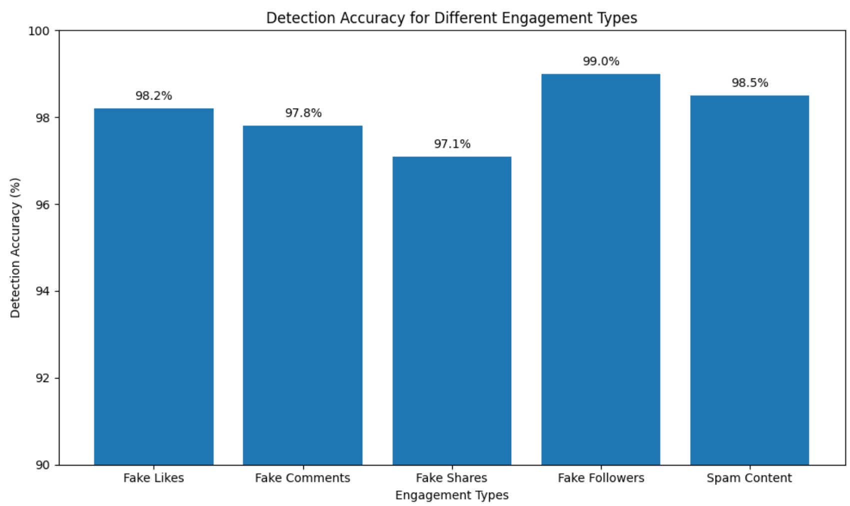

Figure 10: Detection Accuracy for Different Engagement Types

5.3.4 Comparative Analysis of Detection Errors

To compare the error rate of false positive and false negative of the different models, error analysis for detection is performed.

Table 5.18: False Detection Analysis

Algorithm	False Positive Rate (%)	False Negative Rate (%)
Decision Tree	10.4	9.1
Random Forest	5.2	4.8
SVM	7.1	6.5
Logistic Regression	11.3	10.2
LSTM	3.4	2.9
Proposed Hybrid AI Model	1.2	1.0

As seen in the Figure, the proposed model made a significant decrease in the number of detection errors when compared with conventional classification methods.

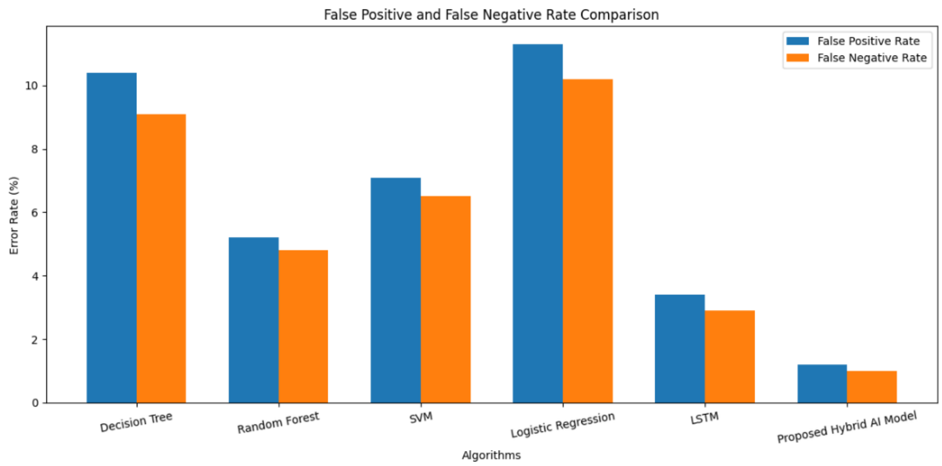

Figure 11: False Positive and False Negative Rate Comparison

6. Conclusion

An extensive analysis of how social media bots impact engagement metrics and the implementation of false engagement strategies was the initial step of the project. In order to comprehend the effects of various forms of fraudulent engagement on digital marketing platforms, we investigated issues such as manipulating followers, making false comments and likes, and distributing misleading content. Based on their ability to mimic human touch and create fake connections, the study also looked into how automated bots play a part in social media analytics and social media marketing decisions. This research introduced a hybrid AI-based detection model that could effectively analyze large-scale social media data. To alleviate the shortcomings of the conventional detection method, this research has proposed a hybrid AI-based detection model that can process the large-scale data of social media effectively. The proposed framework incorporated the various analysis components such as behavioral feature analysis, content-based analysis, NLP and network-based interaction analysis. These methods helped the system to detect more complicated bot activity and coordinated fake engagement activity than traditional methods. Collecting data, cleaning it up, extracting features, training the model, classifying it, and finally, evaluating its performance are all parts of the suggested job. Interactions between bots and actual users were both included in the social media data sets that were evaluated using artificial intelligence and machine learning techniques. Model training was performed with important features from the posts, including posting frequency, engagement rate, follower–following ratio, textual similarity, sentiment patterns, and various types of user interactions. Different ML algorithms, and LSTM networks were implemented and compared during the experimental analysis. The proposed hybrid model of AI is an integration of both DL and behavioral analysis techniques, which is expected to enhance the overall detection capability. The experimental findings showed that the suggested hybrid model outperformed the state-of-the-art ML and DL models. The ROC-AUC values, specificity, and accuracy of the suggested model's categorization were all exceptionally high. In addition, the confusion matrix analysis findings confirmed that the suggested model successfully identified the majority of individuals and bot accounts with minimal misses. With a total accuracy of 98.7%, precision of 98.1%, recall of 97.6%, and an F1 score of 97.8%, the suggested hybrid AI model surpassed the conventional ML algorithms and independent LSTM models in the performance test.

7. Future Scope

It is believed that the precision, scalability, and real-time locating capabilities of the suggested hybrid methodology surpass those of existing, more conventional methods. In addition, the study will show how to improve the accuracy and veracity of engagement metrics on X, Facebook, and Instagram. This study seeks to improve the suggested AI-based architecture in order to deal with the constantly changing and complicated problems that social media ecosystems face. Future research can be done in development of real-time identification algorithms that can detect fraudulent engagement and robots on the social network including X, Instagram and Facebook. If there are coordinated bots that are active on multiple social media platforms at the same time, then there will be more research published regarding how to detect bots on each of the platforms. The use of other types of information such as text, images, videos and interactions with customers can support our deeper analysis of the trends of fraud. Explainable AI can also be used to assist us to solve problems in the future. This would increase availabilities and understanding of the detection models. Questions of ethics regarding data protection, reducing prejudice and following the law are important. Overall, the outcomes could be leveraged to create safer, smarter, and more adaptable digital marketing spaces.

References

[1] Kim, S., & Han, J. (2020, October). Detecting engagement bots on social influencer marketing. In *International Conference on Social Informatics* (pp. 124-136). Cham: Springer International Publishing.

[2] Javed, D., Jhanjhi, N. Z., Khan, N. A., Ray, S. K., Al-Dhaqm, A., & Kebande, V. R. (2025). Identification of spambots and fake followers on social network via interpretable ai-based ML. *IEEE Access*.

[3] Aljabri, M., Zagrouba, R., Shaahid, A., Alnasser, F., Saleh, A., & Alomari, D. M. (2023). ML-based social media bot detection: a comprehensive literature review. *Social Network Analysis and Mining*, *13*(1), 20.

[4] Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: the challenges of AI. *British Journal of Management*, *33*(3), 1238-1253.

[5] Shukla, R., Sinha, A., & Chaudhary, A. (2022). TweezBot: An AI-driven online media bot identification algorithm for Twitter social networks. *Electronics*, *11*(5), 743.

[6] Nevado-Catalán, D., Pastrana, S., Vallina-Rodriguez, N., & Tapiador, J. (2023). An analysis of fake social media engagement services. *Computers & Security*, *124*, 103013.

[7] Akter, N., Abedin, M. Z., Tarafder, M., Nikita, N., Jahan, S., Rimi, N., & Islam, M. (2025). Advanced detection and forecasting of fake news on social media platforms using natural language processing and AI. *Journal of Posthumanism*, *5*(6), 3208-3236.

[8] Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S., & Kaosar, M. (2023). Social media bot detection with DL methods: a systematic review. *Neural Computing and Applications*, *35*(12), 8903-8918.

[9] Jha, A. (2025). How AI Can Prevent Social Media Fraud: A Multimodal Detection Framework. *IJSAT-International Journal on Science and Technology*, *16*(4).

[10] Gannarapu, S., Dawoud, A., Ali, R. S., & Alwan, A. (2020, November). Bot detection using ML algorithms on social media platforms. In *2020 5th international conference on innovative technologies in intelligent systems and industrial applications (CITISIA)* (pp. 1-8). IEEE.

[11] Potluri, R. M., Jumasseitova, A. K., & Potluri, L. S. (2025). The role of AI (AI) in combating digital marketing fraud and bot attacks. In *AI-Enabled Threat Intelligence and Cyber Risk Assessment* (pp. 17-28). CRC Press.

[12] Reddy, D., Singh, A., Chopra, R., & Patel, R. (2021). Leveraging ML algorithms and natural language processing for AI-Enhanced social media marketing analytics. *Journal of AI ML Research*, *10*(8).

[13] Vyas, G., Vyas, P., Muzumdar, P., Chennamaneni, A., Rajavat, A., & Rawat, R. (2024). Extracting and Analyzing Factors to Identify the Malicious Conversational AI Bots on Twitter. *Conversational AI*, 71-83.

[14] Manoharan, A. (2024). Enhancing audience engagement through ai-powered social media automation. *World Journal of Advanced Engineering Technology and Sciences*, *11*(2), 150-157.

[15] Sah, V., Akki, S. N. R., & Shastry, H. K. (2024, December). AI in social media marketing. In *AIP Conference Proceedings* (Vol. 3217, No. 1, p. 020015). AIP Publishing LLC.

[16] Sardana, F., Mishra, K. K., & Tayyab, M. (2024, November). AI-Driven Cyber Security for Brands and Users on Social Media Advertisement. In *2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-5). IEEE.

[17] Usha, S., Kiran, A. U., Almusawi, M. M., & Pravalika, N. (2025, April). Influencer Impact Intelligence Framework (IIIF): An AI-Driven Sentiment and Authenticity Analysis Model for Brand Protection in Social Media Marketing. In *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)* (pp. 1-5). IEEE.

[18] Sadiku, M. N., Ashaolu, T. J., Ajayi-Majebi, A., & Musa, S. M. (2021). AI in social media. *International Journal of Scientific Advances*, *2*(1), 15-20.

[19] Sharma, D., & Jhajj, P. K. (2024). Role of AI as a Monitoring Tool for Social Media Platforms. *Issue 2 Int'l JL Mgmt. & Human.*, *7*, 3351.

[20] Akhtar, M. M., Masood, R., Ikram, M., & Kanhere, S. S. (2024, July). Sok: False information, bots and malicious campaigns: Demystifying elements of social media manipulations. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security* (pp. 1784-1800).

[21] Paschen, J. (2020). Investigating the emotional appeal of fake news using AI and human contributions. *Journal of Product & Brand Management*, *29*(2), 223-233.

[22] Nardello, C., Iannarilli, L., & Giaroli, E. (2025). AI: New challenges, opportunities, and threats in marketing, media, and communication. *Socio-economic Impact of AI: A European Management Perspective*, 97-108.

[23] Shukla, P. K., Veerasamy, B. D., Alduaiji, N., Addula, S. R., Pandey, A., & Shukla, P. K. (2025). Fraudulent account detection in social media using hybrid deep transformer model and hyperparameter optimization. *Scientific Reports*, *15*(1), 38447.

[24] Verma, C., Vijayalakshmi, P., Chaturvedi, N., Umesh, U., Rai, A., & Ahmad, A. Y. B. (2025, February). AI in marketing management: Enhancing customer engagement and personalization. In *2025 International Conference on Pervasive Computational Technologies (ICPCT)* (pp. 397-401). IEEE.

[25] Roy, P. K., & Chahar, S. (2021). Fake profile detection on social networking websites: a comprehensive review. *IEEE Transactions on AI*, *1*(3), 271-285.