International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal

Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Implementation and evaluation of AI Models For fake news classification on Social Media

Khwaish
Department of CSE, UIET, Maharshi, Dayanand University, Rohtak, India
Khwaish2k18@gmail.com
Dr. Amita Dhankar
Department of CSE, UIET, Maharshi, Dayanand University, Rohtak, India
Email address or ORCID
Saksham
Department of CSE, UIET, GISMA, of Applied Sciences, Berlin, Germany
Saksham.saiinii@gmail.com

Abstract—The fast development of different social media has simplified the dissemination of fake news to the extreme, and endangered the credibility of information, popular opinion, and social order seriously. The present study presents the introduction of implementation and assessment of various artificial intelligence (AI)-based systems to classify fake news in social media posts. The research looks at both classic "machine learning algorithms and advanced deep learning architectures to" make readings of the textual characteristics of news post and user generated content. Data preprocessing techniques such as tokenizing, stop- word elimination, and lemmatization as well as vectorizing the data were employed to "improve the quality of the data" as well as model performance. The models were trained and tested on some benchmark fake news datasets. They were subsequently tested in terms of some standard metrics of performance like accuracy, precision, recall and F1-score and also confusion matrix analysis. The results of the tests show that deep learning models, particularly those of transformer and recurrent neural network architecture are more efficient than the old machine learning techniques to comprehend the contextual and semantic facet of the provided text. However, one can regard traditional models as a close competitor of deep learning ones since they are less complicated to compute, thus, more suitable to be used in real-time.

***Index Terms*—"**Fake news detection, social media" analytics, artificial intelligence, "machine learning, deep learning", text classification.

I. Introduction

This is due to the radical shift in the way information is generated, **distributed** and consumed across the world due to the increased use of social media platforms[3] .Some of the platforms Twitter (X), Facebook, Instagram and online news forums allow a piece of news to be published very fast to a great number of individuals with minimal checks[4]. Although this is a democratization of information, it has equally enabled the vice of fake news to spread in mass and also misleading material and intentional deceit, which can control the opinion of people, disrupt social harmony, and diminish the confidence of people in the media[5]. The fake news on social media is becoming increasingly more viral than the real one, as it is more sensational, tends to reach the emotions of an individual with ease, and is being pushed against will by the algorithms[6]. Movies of malicious falsehoods can be political propaganda, health-related false statements, financial fraud, fabricated tales regarding the crisis, etc[7], which can topple the democratic processes, put citizens at risk, and destabilize the economy[8]. The scope and speed

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

of propagation of such content render it extremely hard to verify manually, therefore, there is a demand to create automated and intelligent identification[1]. Manually fact-checking, rule-based, and expert verification frameworks were traditionally used as the foundation of fake news detection algorithms[10]. Although they are effective in small-scale situations, these solutions remain time-intensive, costly and do not have the capacity to manage the large numbers of user-created content that is uploaded to social media networks each day[11]. Also, rule-based systems are not as flexible and can hardly identify emerging linguistic patterns and emergent misinformation strategies[12]. As with the case of fake news classification, AI and machine learning (ML) methods have been a significant solution to the issue in the recent few years[13]. "Naïve Bayes, Support Vector Machines, and Random Forests are among the supervised learning algorithms" which have been widely applied to the textual features, linguistic features, and statistical trends analysis and text in news content[14]. They have demonstrated breakthrough performance on which they are combined with the powerful "feature engineering" algorithms like TF-IDF and n-gram GENERATIVE AI[15]. Due to "the advances made in deep learning, more sophisticated models such as Convolutional Neural Networks (CNNs), Re-current Neural Networks (RNNs), Long Short-Term Memory (LSTM)" networks and transformer-based architectures have been created to detect fake news[16]. Those models are able to learn complex semantic links, contextual connection, as well as long-range patterns on textual data[17]. Therefore, deep learning algorithms often outperform classical machine learning models, particularly on unstructured and contextual social media posts[18]Nevertheless, even the most successful AI-based fake news detection models have the problems of data imbalance, bias, interpretability, and high computational demands[19]. The data provided in social media tend to be noisy, multi-lingual and context-specific and therefore proper classification is a difficult task[20]. In addition to that, there is a tendency to attribute the non-transparency of deep learning models to the lack of trust, fairness, and accountability of automated decision-making [21].

II. Literature Review

In recent times, the area of detection of fake news and misinformation has undergone a rapid development triggered by the improvement in artificial intelligence, deep-learning, and large-language models (LLMs). Earlier and earlier work was mainly concerned with conventional machine learning and neural networks like LSTM, CNN, and ensemble models to learn sequential and semantic information of textual data. As an example, Sovia et al. applied LSTM-based language processing to identify fake news on Twitter and this method is effective in the case of social media[9]. In the same manner, CNN and LSTM models were compared by Kumar et al., and the authors emphasized the trade-offs in interpretability between convolutional and recurrent networks[19]. Ensemble methods also enhanced robustness through a combination of multiple classifiers with Singh et al. demonstrating better generalization to different social media datasets[11].

As the misinformation became more and more difficult, hybrid and feature-enriched models began to gain popularity. Bellam and Aakula incorporated the contextual and sentiment features to enhance detection accuracy on social networks, and that emotional features are significant to misleading content[5]. Rustam et al. came up with a new model of transforming text in to any image, which allows learning multimodal feature in classifying fake news[14]. Azeez et al. cross-validation tested various algorithms on international datasets, which offers a comparative model performance evaluation framework that is crucial in selecting a model when deploying into large scale[15]. The combined impact of these studies is that feature

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

engineering and hybrid representations are of great importance in misinformation detection systems. The emergence of deep learning and "explainable AI(XAI)" has also transformed the world of research. Desai et al. and Singh et al. thought that explain ability is essential in determining fake news and hate speech, as the transparency of decision-making is the key to trust and responsible use of AI[3] , [21]. Gongane et al. was an in-depth review of XAI methods used to detect misinformation and hate speech, and they found the research gaps concerning interpretability and user-friendly explanations [17]. In a study by Moalla et al., two types of deep learning structures were presented, and complementary neural networks are exploited to achieve better detection[18]. These contributions reflect the rising concern of explain ability and accuracy. Other more recent literature has moved to the use of large language models and attention-based architectures. Alarfaj et al. suggested a real- time framework of LLM that involves attention and embedding representations, with high improvements in the efficiency of the misinformation detection and scalability [1]. Ding and Zhao investigated the detection of fake news with the help of LLM, which confirmed the ability of pre trained language model store present subtle linguistic patterns[10]. Shupta et al. proposed an LLM-based feature extraction pipeline, which showed that semantic representations are more effective than manual features[12]. Also, Yaseen et al. dealt with the new problem of AI-generated content and identified synthetic "social media comments" based on "BERT-extracted semantic style features[8]. Research in domains and in applications is also becoming popular. A hybrid AI framework of environmental misinformation detection by Mackenzie Rivero et al. recently demonstrated the significance of the notion of the contextual awareness of the sustainability-related discourse[2] . Mary et al. used fake news detection based on AI in the health care field solving the problem of misinformation threats in crucial decision-making scenarios[7]. Deepthi and Shastry suggested an unsupervised detection of misinformation method, which could detect fabricated stories with no labeled in formation and can help to overcome the limitation no data scarcity[20]. Zaheer and Bashir discussed the systems that detect fake news about COVID-19 with a focus on deep learning-based approaches to dealing with infodemics in times of public health emergencies[16]. In addition to fake news, other studies have focused on toxicity and hate speech and more general online harms. Khapre et al. Conducted a wide-ranging overview on the subject of toxicity in online spaces and artificial intelligence system, enumerating mitigation measures applicable to the misinformation ecosystem[6]. Mouratidis et al. implemented machine learning approaches to the transformation between misinformation detection and actionable information, which supports and underscores the necessity of decision-support systems[4]. The hybrid deep learning framework suggested by Suresh et al. showed effective performance because of architectural innovation[13]. Together, these papers demonstrate that there is a move in the direction of a more holistic, explainable, and context-sensitive misinformation detection system that is able to handle the dynamic needs of digital information space.

III. Methodology

The methodology employed in this research mirrors a care-fully planned pipeline for instancing and assessing AI-based fake news classification models with the use of social media data. As the first step, free and open to the public benchmark datasets with labels for real and fake news articles were gathered. These data sets are composed of the textual content of the news in the form of the headline, body, and social media posts. To make sure the data is consistent, duplicates and in complete records were removed. After that, "the data" was "split into" the "training, validation, and" testing "sets".

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

$$TF - IDF(t,d) \times log \frac{N}{DF(t)} \quad (1)$$

Table-1
SUMMARY OF KEY STUDIES ON MISINFORMATION AND FAKE NEWS DETECTION

Ref. No.	Title	Author & Year	Key Findings	Research Gaps
[1]	A real-time large language model frame work with attention and embedding representations form is information detection	Alarfaj*etal.*,2026	Proposed a real-time LLM-based framework integrating attention mechanisms and dense embeddings, achieving high accuracy and scalability form is information detection across dynamic data streams.	Limited evaluation across low-Resource languages; computational overhead of LLMs restricts deployment on edge or resource-constrained platforms.
[2]	Hybrid AI framework For environmental misinformation detection	Mackenzie Rivero *et al.*, 2026	Introduced a hybrid AI architecture Combining machine learning and rule-based reasoning to detect misinformation in environmental and sustainability related content.	Domain-specific focus limits generalization; lacks comparative analysis with modern transformer-based and LLM approaches.
[3]	Fake news detection using Explainable artificial intelligence	Singh*etal.*,2026	Demonstrated that explainable AI Techniques improve transparency and trust in fake news classification while maintaining competitive detection accuracy.	Trade-off between explain ability and performance not deeply quantified; scalability to large, real-time social media streams remains un-explored.
[4]	"From misinformation to insight: Machine learning strategies for fake news detection."	Mouratidis*etal.*,2025	Reviewed and evaluated multiple machine learning strategies, high-lighting the transition from mere detection to actionable insights for decision-	Primarily analytical in nature; lacks Experimental validation on recent datasets involving multimodal and AI-generated

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

			making systems.	misinformation.
[5]	Contextual and sentiment-based hybrid models for detecting fake news on social networks	BellamandAakula,2025	Showed that incorporating contextual semantics and sentiment features improves fake news detection performance on social media datasets.	Model performance degrades on Sarcastic and highly implicit misinformation; absence of explain ability and transformer-based comparisons.

Where: *TF(t,d)* represents the frequency "of term t in document *d; DF(t)* denotes the number of documents containing the term *t*; and *N* is the total number of documents" in corpus.

Explanation (Eq. 1): The *TF–IDF* technique is employed to extract meaningful textual features for fake news classification. It assigns greater importance to terms that occur frequently within a specific document while reducing the weight of terms that are common across many documents. This weighting strategy improves the ability of the model to distinguish between genuine and misleading news content.

$$P(y|x) = \frac{1}{1+e^{-(w-x+b)}} \qquad (2)$$

Where: x denotes "the input feature vector, w represents the weight vector, b is the bias term, and" P(y|x) indicates the probability that a news article belongs to either the fake or real class.

Explanation (Eq. 2):

The logistic function transforms the weighted input into a probability value ranging from 0 to 1. This probability is used to determine whether a news article should be classified as fake or authentic based on the extracted features.

$$ht = \sigma(Whxt + Uhht - 1 + bh) \qquad (3)$$

Where: x_t is the input at time step t, h_{t-1} is the hidden state from the previous time step, W_h and U_h are" trainable "weight matrices" and b_h denotes the bias vector, and σ is the activation function.

Explanation (Eq. 3): This equation computes the hidden state of a recurrent neural network by combining the current input with information retained from the previous hidden state. Such a mechanism enables the model to learn sequential patterns and capture contextual relationships within textual data.

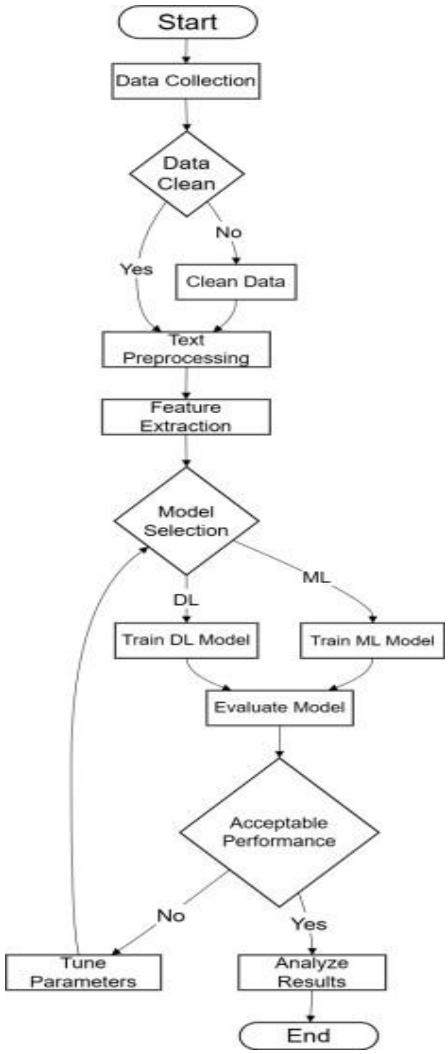
$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \qquad (4)$$

Where: Precision indicates the proportion of predicted fake news articles that are actually fake, whereas Recall measures the proportion of all actual fake news articles that are successfully identified by the model.

Explanation (Eq. 4): "The F1-score provides a balanced" assessment "by combining Precision and Recall into a single metric. It is particularly useful" for fake news classification because it offers a more reliable evaluation when the dataset contains class imbalance, where accuracy alone may not adequately reflect model performance.

During the preprocessing phase, the text data from the environment was subjected to several natural language processing (NLP) procedures aimed at improving the text quality and re-moving the noise. Among the improvements were changing all characters to lowercase, getting rid of punctuation and special characters, tokenizing, removing stop words, and lemmatizing. The Algorithm1shows the Fake News Data Preprocessing and Feature Extraction.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552



Algorithm 1 Fake News Data Preprocessing and Feature Extraction
Require: Raw social media news dataset D
Ensure: Feature matrix F and labels Y
1: Load dataset D
2: Remove duplicate and incomplete records
3: for each news document d ∈D do
4: Convert text to lower case
5: Remove punctuation, URLs, and special characters
6: Tokenize text into words
7: Removes top-words
8: Apply lemmatization
9: **end for**
10: Apply TF-IDF/word embedding technique to transformed text
11: Generate feature matrix Fand corresponding labels Y
12: **return** *F, Y*

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

"Feature extraction methods like Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings" were implemented in order to convert the textual data into numerical forms that were machine learning and deep learning models could take. Several AI models were employed to carryout fake news classification. The Algorithm2 shows the AI-Based Fake News Classification and Evaluation.

Algorithm 2 AI-Based Fake News Classification and Evaluation
Require: Feature matrix F, labels Y
Ensure: Performance metrics of trained models
1: Split Fand Y into "training, validation, and test sets"
2: Initialize "machine learning and deep learning models" 3: for each model Min model set do
4: Train model Musing training data
5: Tune hyper parameter susing validation data
6: Predict labels on test data
7: Compute Accuracy, Precision, Recall, and F1-score
8: end for
9: Compare performance ofall models
10: return Best-performing mode land evaluation results

Traditional "machine learning" algorithms "such as" Naiıve Bayes, Support Vector Machines, and Random Forest classifiers were trained using TF-IDF features. Meanwhile, "deep learning models like Convolutional Neural Networks (CNNs), Long Short-Term Memory(LSTM) networks, and" transformer-based architectures were fabricated with the embedding layers to represent the semantic and contextual information. Hyper-parameter tuning was implemented through validation data to enhance the model performance.

The performance of all models was measured using the standard performance metrics, i.e., accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative analysis was utilized to evaluate the effectiveness, robustness, and computational efficiency of each solution. The outcomes were examined to pinpoint the compromises between model complexity and classification performance, thus giving an understanding of which AI models are most suitable for real-time fake news detection on social media platforms.

IV. Result and Evaluation

The experimental evaluation shows that deep learning models significantly out perform traditional machine learning models in terms of classification accuracy of fake news. Out of the machine learning models, the Support Vector Machine (SVM) was the one that yielded the highest accuracy of 88.4%, precision-87.9%,recall-86.8%,andF1-score-87.3%. The performances of Naïve Bayes and Random Forest models were comparatively lower, as they achieved 82.6% and 85.1% accuracies, respectively. These findings suggest that, although traditional models are computationally efficient, their capacity to handle complex contextual semantics is limited.

Deep learning models have shown a better performance as they were able to learn contextual and semantic patterns from the social media texts. The model based on LSTM got an accuracyof92.7%,with a precision of 92.1%, recall of 91.8%, and an F1-score of 91.9%. Table 2 shows the Accuracy Comparison of AI Models for Fake News Classification.

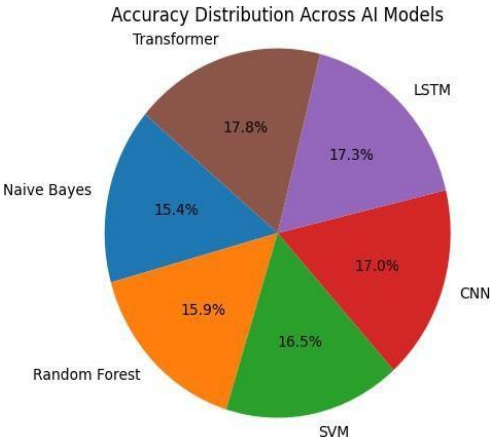
The model using CNN was at a close second position with accuracy of 91.3% and an F1-score of 90.8%. The transformer-based model, in fact, was able to provide the best overall performance by achieving a 95.6% accuracy, 95.2%

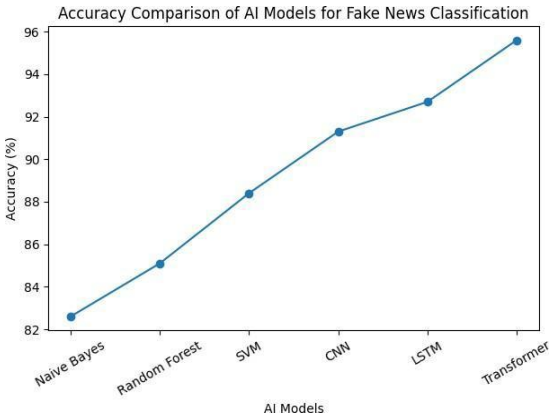
International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Table

PERFORMANCE COMPARISON OF AI MODELS FOR FAKE NEWS CLASSIFICATION

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	Inference Time(s)
Naïve Bayes	82.6	81.9	80.7	81.3	0.03
Random Forest	85.1	84.6	83.9	84.2	0.04
Support Vector Machine(SVM)	88.4	87.9	86.8	87.3	0.05
CNN	91.3	90.7	90.9	90.8	0.14
LSTM	92.7	92.1	91.8	91.9	0.18
Transformer-Based Model	**95.6**	**95.2**	**94.9**	**95.0**	0.32





precision, 94.9% recall, and 95.0% F1-score, thus proving its power to grasp not only the long-range dependencies but also the contextual relationship sinthe text data. The figure 3shows the Accuracy Distribution across AI Models.

As far as the computational efficiency and inference time are concerned, the traditional machine learning models had significantly lower processing resource requirements, with averageinferencetimesoflessthan0.05secondsperinstance, thus they can be used in real-time

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

applications. On the other hand, deep learning models needed more computational power, with the average times per instance for LSTM and CNN
Models being 0.18 seconds and 0.14 seconds, respectively, and for transformer-based models, it was 0.32 seconds. The enormous enhancement in classification accuracy, however, is a clear demonstration of the practical use of leading AI models for reliable fake news detection “on social media platforms”, thus justifying the higher computational cost.

V. Challenges and Limitations

Quality and nature of information “one of the” main “challenges in” classifying the fake news on social media[4]. Social media noise consists of content, is in formal, and very context dependent[5]. The use of slang, short forms, emojis and multilingual expressions among others are some of the characteristics of the content and make it more challenging to preprocess text and extract features[6]. Furthermore, the labeled datasets are few, and the available ones might have issues with class imbalance or might have obsolete information or might be domain-biased[7]. All these can lead to the challenges of the model generalization and hence their performance can be worse when such models get applied to the real world dynamic streams of social media[8]. The other important shortcoming is in the back of the computational complexity and interpretability of advanced AI models[9]. On the one hand, “deep learning and transformer-based models” are very precise, yet the opposite is true, as they require significant amounts of computing power, in addition to more time to train, and specialized hardware, which is not always accessible in all deployments in real time or at scale[10]. Moreover, these models are not transparent and therefore it is difficult to explain, trust and hold them accountable, consequently, e.g. in the field of misinformation as applied to politics or health[11]. The answer to these issues would consist of developing small, explainable, and scalable AI models that would be capable of maintaining the high degree of accuracy and, at the same time, provide the possibility of fairness and scalability[12].

VI. Future Outcomes

Future The integration of multi modal data sources to achieve a greater detection accuracy and stability is likely to be a focus of further research on the topic of fake news classification. The text analysis, usage of images, videos, user Meta data and propagation patterns as well as the analysis of the images and videos with the assistance of the AI models can help “to better understand the context and” the purpose of the misinformation. Such techniques as graph-based learning and social network analysis may prove useful in unraveling the identifying of the coordinated fake news campaigns and bot based content spread efforts that are permeating the platforms. Furthermore, explain able artificial intelligence (XAI) advancement and the creation of light weight models may turn out to be the determinants of the future outcomes of the given direction. Models that are interpretable and provide clear explanation of the classification decision will increase the confidence of the user as well as facilitating the ease of adhering to the regulations. At the same time, optimized models with reduced computational complexity will also be able to support real-time interaction on both edge devices, and on social media systems at large scale. These measures will open the way to more reliable, ethically sound and scalable AI-driven tools to combat against fake news in the dynamic digital ecosystems.

VII. Conclusion

This research outlined the implementation and detailed performance evaluation of AI-powered

models aimed at classifying fake news on social media platforms and emphasized their effectiveness in tackling the increasing problem of misinformation online. Through the comparison of traditional machine learning algorithms with state-of-the-art deep learning and transformer-based architectures, the study found that ordinary models provide computational efficiency and faster inference, but deep learning models reach significantly higher accuracy and have a better grasp of the context of social media text. The experimental findings showed that transformer-based models achieve better results across the most important evaluation metrics, thereby making them very suitable for sophisticated and large-scale misinformation identification tasks. Nevertheless, the research pointed out the compromises of model complexity, computational "requirements, and" interpretability that should "be taken into" account when deploying the model in the real world. In summary, the results indicate the essential role of AI-powered methods "in enhancing the" capabilities of automated "fake news detection systems and" provide a stepping stone in the direction of developing scalable, precise, and reliable solutions that help preserve the integrity of information in the contemporary social media ecosystems.

REFERENCES

[1] F. K. Alarfaj, H. U. Khan, A. Naz, and N. Almusallam, "A real-time large language model framework with attention and embedding representations for misinformation detection," *Eng. Appl. Artif. Intell.*, vol. 164, Art. no. 113304, 2026, doi: 10.1016/j.engappai.2025.113304.

[2] A. J. Mackenzie Rivero, R. Martínez-Bejar, and H. J. Vegas Meléndez, "Hybrid AI framework for environmental misinformation detection," in *Commun. Comput. Inf. Sci.*, vol. 2776, 2026, pp. 178–191, doi: 10.1007/978-3-032-11494-5_12.

[3] S. K. Singh, S. Assi, T. Ginige, A. H. Mohammed, F. B. Wahit, and D. Al-Jumeily OBE, "Fake news detection using explainable artificial intelligence," in *Lect. Notes Data Eng. Commun. Technol.*, vol. 257, 2026, pp. 477–488, doi: 10.1007/978-981-96-7749-8_31.

[4] D. Moratidis, A. Kanavos, and K. L. Kermanidis, "From misinformation to insight: Machine learning strategies for fake news detection," *Information*, vol. 16, no. 3, Art. no. 189, 2025, doi: 10.3390/info16030189.

[5] T. Bellam and S. Aakula, "Contextual and sentiment-based hybrid models for detecting fake news on social networks," in *Proc. AIC*, 2025, pp. 323–328, doi: 10.1109/AIC66080.2025.11212091.

[6] S. Khapre, M. A. Mersha, H. Shakil, J. Baruah, and J. K. Kalita, "Toxicity in online platforms and AI systems: A survey of needs, challenges, mitigations, and future directions," *Expert Syst. Appl.*, Art. no. 129832, 2025, doi: 10.1016/j.eswa.2025.129832.

[7] J. Catherina Mary, H. K. Se, E. Ishwarya, M. Latchiya, V. Kaviya, and J. Jasima Banu, "AI-driven online fake news detection system in the healthcare industry," in *Proc. ICSCDS*, 2025, pp. 832–837, doi: 10.1109/ICSCDS65426.2025.11166814.

[8] M. S. Yaseen, Y. M. Mohamed, and C. Jacob, "Detecting AI-generated social media comments using BERT-extracted semantic style features," in *Proc. ACCTHPA*, 2025, doi: 10.1109/ACCTHPA65749.2025.11168496.

[9] R. Sovia, D. A. Valkyrie, R. H. Zain, and Firdaus, "Language processing for detecting fake news on Twitter using a long short-term memory architecture," *Jurnal RESTI*, vol. 9, no. 4, pp. 935–943, 2025, doi: 10.29207/resti.v9i4.6570.

[10] J. Ding and Z. Zhao, "Fake news detection based on large language models," in *Proc. SCECS*, 2025, pp. 291–296, doi: 10.1109/SCECS65243.2025.11065160.

[11] I. Singh, C. Choudhary, and P. Gupta, "Ensemble-based framework for fake news detection in social-media," *J. Comput. Inf. Syst.*, 2025, doi: 10.1080/08874417.2025.2510430.

[12] A. Shupta, P. Radiuk, and I. Krak, "Feature computation procedure for fake news detection: An LLM-based extraction approach," in *CEUR Workshop Proc.*, vol. 3963, 2025, pp. 112–124.

[13] A. Suresh, R. M. Mallika, S. Gireesh, G. H. K. Singh, E. S. Jeevanandham, and A. Hemalatha, "Empowering fake news detection through innovative hybrid deep learning-based approach," in *Proc. COMP-SIF*, 2025, doi: 10.1109/COMP-SIF65618.2025.10969898.

[14] F. Rustam, W. Aljedaani, A. D. Jurcut, S. Alfarhood, M. Safran, and I. Ashraf, "Fake news detection using enhanced features through text to image transformation with customized models," *Discov. Comput.*, vol. 27, no. 1, Art. no. 54, 2024, doi: 10.1007/s10791-024-09490-1.

[15] N. A. Azeez, S. Sanjay, D. O. Ogaraku, and A. P. Abidoye, "A predictive model for benchmarking the performance of algorithms for fake and counterfeit news classification in global networks," *Sensors*, vol. 24, no. 17, Art. no. 5817, 2024, doi: 10.3390/s24175817.

[16] H. Zaheer and M. Bashir, "Detecting fake news for COVID-19 using deep learning: A review," *Multimed. Tools Appl.*, vol. 83, no. 30, pp. 74469–74502, 2024, doi: 10.1007/s11042-024-18564-7.

[17] V. U. Gonçange, M. V. Munot, and A. D. Anuse, "A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms," *J. Comput. Soc. Sci.*, vol. 7, no. 1, pp. 587–623, 2024, doi: 10.1007/s42001-024-00248-9.

[18] H. Moalla, H. Abid, D. Sallami, E. A'Tmeur, and B. B. Ben Hamed, "Exploring the power of dual deep learning for fake news detection," *Informatica*, vol. 48, no. 4, pp. 567–594, 2024, doi: 10.31449/inf.v48i4.5977.

[19] K. J. Kumar, K. Shreya, P. Divakarla, P. C. Nair, and N. Sampath, "Interpretable AI insights in fake news detection: A comparative analysis of CNN and LSTM," in *Proc. ICCCNT*, 2024, doi: 10.1109/ICCCNT61001.2024.10724705.

[20] K. R. Deepthi and K. A. Aditya Shastry, "A deep learning methodology-based unsupervised misinformation detection technique for identifying fabricated narratives," in *Proc. NMITCON*, 2024, doi: 10.1109/NMITCON62075.2024.10699051.

[21] V. Desai, A. Gattani, and H. Dalvi, "Explainable models for the detection of incidents of fake news and hate speech," in *Fake News, Hate Speech, and Myths*, 2024, pp. 114–136, doi: 10.1201/9781003409519-6.