iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71831

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

# The Role of AI in Social Media Misinformation: Strategies and Impacts

Akash Kumar[1], Asst. Prof. Shyam Shankar Dwivedi[2]
*AIIT, Amity University Patna*

***Abstract: Artificial intelligence (AI) has significantly influenced content distribution on social media, but it has also contributed to the rapid dissemination of misinformation. AI-powered technologies, such as deep fake manipulation, automated content creation, and engagement-driven algorithms, facilitate the swift production and amplification of false information. This paper investigates how AI is leveraged to spread misinformation across major social media platforms, including Facebook, Twitter (X), YouTube, and Instagram. Through a case study approach, we examine real-world incidents where AI-generated content has misled users, particularly in political campaigns, public health crises, and viral digital trends. Social media algorithms, designed to boost engagement, often unintentionally amplify misleading content, making the detection and regulation of misinformation more challenging. This study explores the difficulties in identifying AI-generated fake news and assesses the effectiveness of current fact-checking mechanisms and moderation strategies. While AI plays a role in propagating misinformation, it can also serve as a solution through advanced natural language processing (NLP), deep fake detection technologies, and automated verification systems. The findings emphasize the necessity of ethical AI usage, enhanced content moderation techniques, and stricter regulatory frameworks to curb AI-driven misinformation. Strengthening detection technologies and raising public awareness can significantly reduce the impact of false information on social media ecosystems.***
***Keywords: AI misinformation, social media algorithms, deep fake detection, fake news regulation.***

## I. INTRODUCTION

The rise of social media has significantly changed the way people interact with news and information. Rather than depending solely on newspapers or television, individuals now rely heavily on platforms like Facebook, X (formerly Twitter), and Instagram to stay informed and share their views. While this shift has made communication faster and more inclusive, it has also made it easier for false and misleading information to spread widely. Misinformation has emerged as a serious issue, influencing public behavior, polarizing opinions, and even disrupting democratic processes [1]. Artificial Intelligence (AI) plays a crucial but complicated role in this environment. On the negative side, AI-powered algorithms that prioritize user engagement often end up promoting sensational or emotionally-driven content—sometimes regardless of its truthfulness [2]. Additionally, advances in generative AI, such as deep fake videos and AI-written articles, have made it easier to produce deceptive content that appears credible. These technologies have been used to manipulate public opinion on political issues, health topics, and more [3]. However, AI is also proving to be a powerful tool in identifying and controlling misinformation. With technologies like machine learning, natural language processing (NLP), and image recognition, AI systems can detect fake news, flag harmful content, and recognize automated bot activity [4]. These tools are already in use by major tech companies to monitor platforms and reduce the spread of false information. Despite these advancements, challenges remain. AI doesn't always understand nuance, context, or sarcasm, which can lead to mistakes—either by mislabeling genuine content or failing to detect harmful posts. Ethical concerns also arise around privacy, free speech, and the fairness of automated decision-making [5]. This research paper explores both the harmful and helpful sides of AI in relation to misinformation on social media. It aims to analyze how AI contributes to the problem, the strategies being used to mitigate its effects, and the broader societal and ethical consequences. The goal is to understand how AI can be used more effectively and responsibly to support a safer and more truthful digital space. As technology progresses, the ability to detect misinformation becomes more complicated and thus more difficult to detect using standard machine learning (ML) techniques. This motivates our focus on deep learning (DL) techniques for the problem of fake news detection. In this systematic literature review (SLR), we investigate existing fake news detection (FND) strategies that use deep learning. We focus on publicly available datasets used in FND and their NLP approaches [6]. We aim to gather information about the transfer learning techniques applied and the methods used for addressing class imbalance, to examine their effect on detection accuracy. Our survey aims to identify open issues and research gaps in current studies. To the best of our knowledge, we are the first to provide a comprehensive SLR that investigates the effects of transfer learning and class imbalance treatment in the fake news detection domain [7].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

## II. LITERATURE REVIEW: FAKE NEWS DETECTION USING MACHINE LEARNING AND DEEP LEARNING APPROACHES

### A. Search Strategy Overview

The rapid spread of misinformation on social media platforms has sparked significant research interest in fake news detection. This literature review brings together key methods, challenges, and recent innovations in using machine learning (ML) and deep learning (DL) for this purpose. The widespread circulation of false information on platforms like Facebook and Twitter has led to the development of automated systems for detecting fake news [8]. These systems are essential for maintaining the integrity of information shared online. While social media has become a primary source of news for many, it is often exploited by malicious actors who spread false content for financial gain or political manipulation. Researchers studying lowresource languages such as Arabic have also pointed out unique challenges, including limited datasets and language-specific complexities [9]. Several studies highlight the strengths of DL over traditional ML methods. Earlier works focused mainly on data mining and basic ML techniques. More recent research explores advanced DL models like Generative Adversarial Networks (GANs), Bidirectional Encoder Representations from Transformers (BERT), and Attention mechanisms, which have shown superior capabilities in capturing linguistic and contextual patterns [10]. DL models have also been particularly effective in dealing with challenges like class imbalance and transfer learning, especially during highimpact events like elections or the COVID-19 pandemic [11]. One notable contribution is the introduction of optimized models like OPCNN-FAKE, which have outperformed conventional ML models such as Decision Trees, Logistic Regression, K-Nearest Neighbor, Random Forest, Support Vector Machine, and Naive Bayes. These models use techniques like N-gram, TF-IDF, and GloVeembeddings and have demonstrated strong performance across benchmark datasets [12], [13]. Other researchers have focused on the power of feature engineering to improve classification results. By analyzing features related to news sources, publication patterns, and user interaction on social platforms, these studies provide valuable insights into how false news spreads [14]. In particular, some studies have shown that SVM-based models in Indian contexts can detect fake news with up to 93.6% accuracy [15]. Ensemble methods that combine bagging and boosting algorithms—specifically Random Forest and CatBoost—have also shown exceptional accuracy levels of 99% and 98%, respectively. These methods are among the most robust approaches for handling misinformation. In addition to model development, some studies explore the complexities of determining the credibility of content and its sources. They argue that DL offers more comprehensive detection mechanisms compared to traditional ML [16]. Finally, AI tools, including both ML and DL, are being applied across fields like politics, business, and sports. These tools not only show high accuracy in fake news detection but also offer a foundation for future research aimed at refining and scaling these systems.

### B. Research Questions

The key focus of our SLR is on understanding how the DL techniques have been used to address the FND problem.
RQ1: What is fake news and how it does affect people and society?
RQ2: Which deep learning algorithms have been used for fake news detection throughout time?
RQ3: How effective are deep learning methods for fake news detection?
RQ4: How to create a pertinent model for detecting fake news? RQ5: How to prevent the spread of fake news?

### C. Source Databases And Search Query

Our literature search was conducted across multiple reputable academic databases, including IEEE Xplore, ACM Digital Library and Google Scholar [17]These platforms were selected due to their extensive collections of peer-reviewed journals, conference papers, and scholarly articles in the fields of computer science, artificial intelligence, and media studies.
The search queries were constructed using combinations of keywords and Boolean operators such as:
- “fake news” AND “deep learning”
- “AI for misinformation detection”
- “social media” AND “fake content prevention”
- “misinformation” AND “natural language processing”

The search was refined by applying filters such as publication year (2015–2024), language (English), and document type (journal articles, conference papers, and review articles).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue V May 2025- Available at www.ijraset.com*

## III. DEEP LEARNING ALGORITHMS USED FOR FAKE NEWS DETECTION

Looking closely at different models and techniques shows just how important deep learning (DL) has become in handling tasks like fake news detection. This became even more urgent during the COVID-19 pandemic, when misinformation and rumors were spreading rapidly across social media. As a result, developing smarter and more reliable algorithms wasn't just useful—it became a real necessity. Figure 1 demonstrates a clear increase in the use of DL models over the years[18],[19],[20],[21].

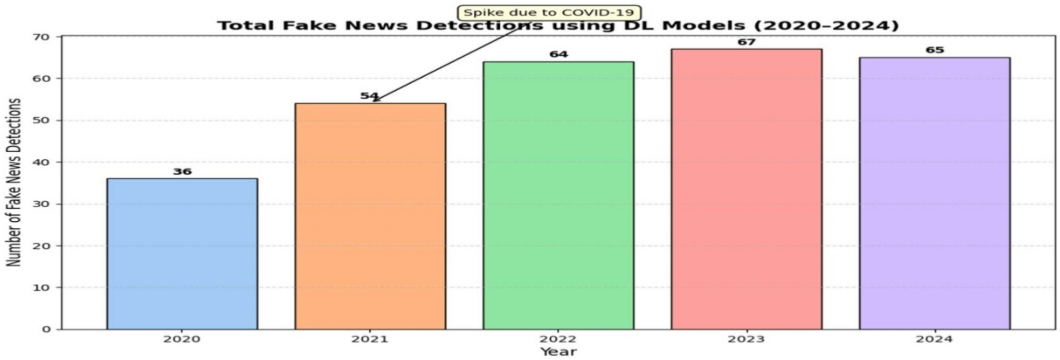
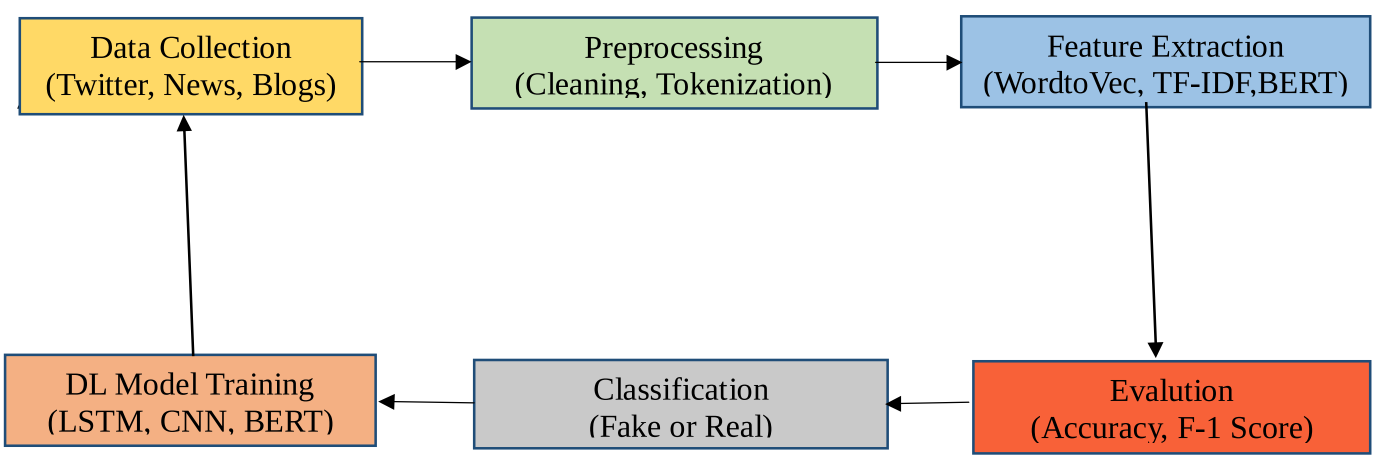



Figure 1. DL Models used for FND between the years 2020 and 2024.

From the data reviewed, it's clear that fake news detection (FND) usually follows a fairly common process, shown in Figure 2. It typically starts with collecting or building a dataset—most researchers use publicly available news articles for this step. After that, the data is cleaned and prepared using preprocessing techniques so it can be fed into a neural network [22],[23] To turn text into a format that machines can understand, methods like Word2Vec and GloVe are often used to convert words into numerical vectors [24]. Once everything is ready, the neural network is trained, and predictions can be made [25]. Neural networks used in FND come in different forms, depending on how they're built and how they handle data. The simplest are feedforward networks, like single-layer or multi-layer perceptron. Then there are convolutional neural networks (CNNs), which are especially good at working with grid-like data such as images. These include basic CNNs as well as more complex models like residual and dense networks. Based on the data collected from the surveyed articles, it's clear that researchers have widely explored various deep learning (DL) algorithms for fake news detection. Figure 6 illustrates how often different DL models were used across the reviewed studies. Specifically, it shows the percentage of papers that used each model. Among them, the (Bi) LSTM model was the most popular, appearing in 72% of the articles. CNNs followed closely, being used in 61% of the studies. The third most common approach was hybrid architectures, which combine different types of neural networks to improve performance [27],[28].

It's worth noting that many studies used more than one model, so the percentages in Figure 6 add up to more than 100%. The next sections take a closer look at these major architectures and how they are applied in fake news detection.



### *A. Architectures Based On Convolutional Neural Networks*

Our research demonstrates in Figure 3 that around 61% of earlier researches involve CNN-based systems when considering the FND task, whether CNN-based models are built as individual systems or as a part of more complex architectures [29],[30],[31]. This common use case demonstrates CNN's ability to learn hierarchical patterns and features from the word to the sentence level on textual data, which is important when picking up subtle cues -- a common trait of misinformation.

In this work, we use a CNN-based model designed specifically for fake news detection. The presented model is composed of several crucial layers to cooperatively extract and transform input features for classification. The architecture starts with an input layer that pads text data to have a length of 1000 words. This normalization makes the descriptions of different lengths consistent with each other to some extent and helps them to be processed efficiently in the network[32],[33].

After the input layer, an embedding layer is used to transform each token into a dense vector. This layer is important for dimensionality reduction/binding of semantic word relationships. In our case, embedding dimension is selected as 100 for trade-off between computational affordability and representational power. The main structure of the architecture is composed of three Conv-channel pools. All convolutional layer in these blocks have 128 5 5 filters with ReLU activation, which is used to add non-linearity and avoid vanishing gradient problem. These convolutional layers search the input embeddings for local patterns that signal fake or true content, such as particular word combinations or syntactic structures.

Following the convolutional blocks, the output feature maps are flattened with a flatten layer to one dimensional one. A subsequent dense (fully connected) layer enables the extracted intensity pattern to be combined over the entire input. Finally, a softmax-activated output-layer is applied for classification, where it outputs a probability distribution over to the target classes (e.g., fake or real news).The architecture is designed to trade-off the required depth and the computational complexity in order to learn complex features without overfitting. Furthermore, the modular design of CNN architecture makes it easy to insert in future hybrid models the other components, such as attention mechanism or recurrent layers, which can potentially further improve task of detecting fake news.

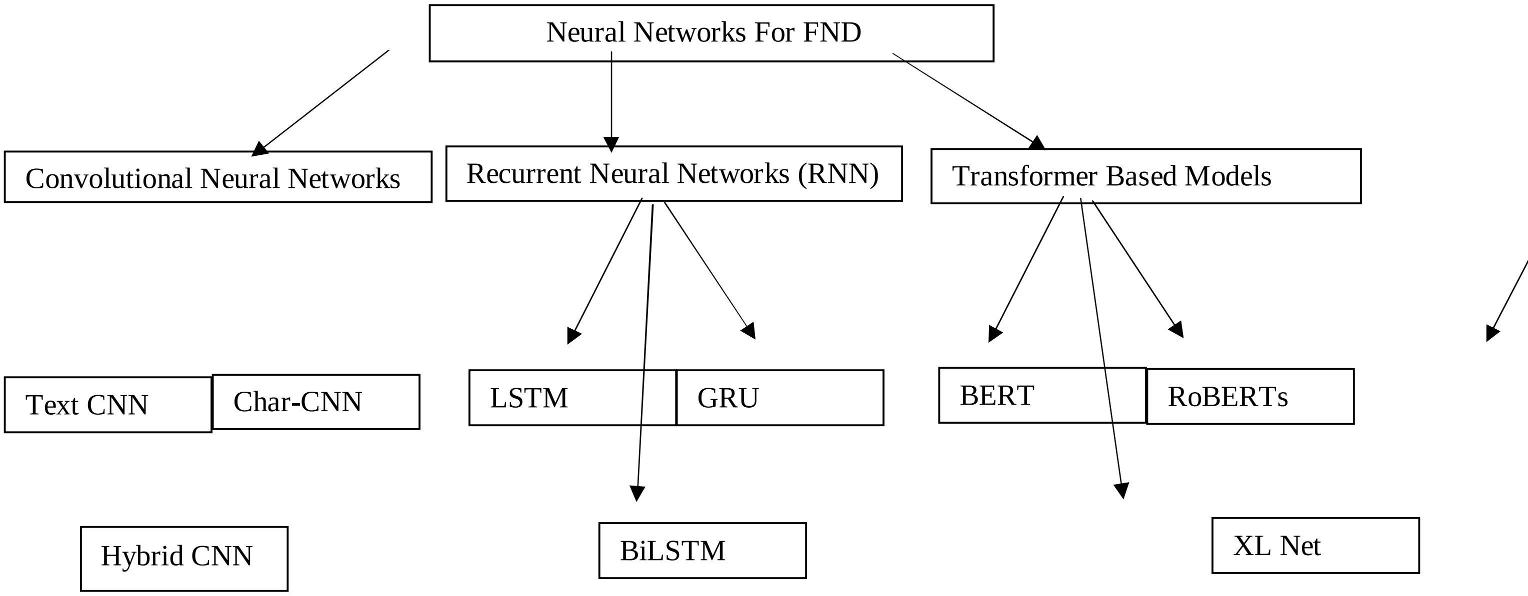



Figure 3. Taxonomy of Neural Network Architectures.

### *B. Architectures Based On Recurrent Neural Networks*

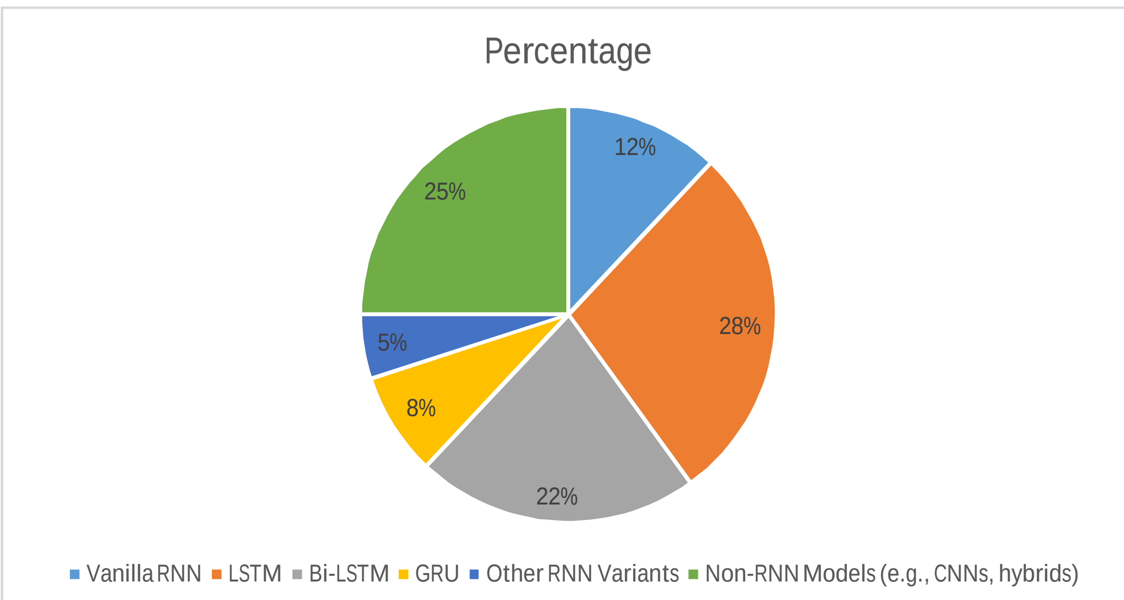
Another most exploited category of algorithms in prior work in FND task is the Recurrent Neural Network (RNN) and its several derivatives. Figure 4. Shows the RNNs are well-suited to sequential data and are therefore a natural choice for text analysis. vanilla RNN) we experiment on several RNN-based architectures, including LSTM, GRU, unidirectional LSTM and Bidirectional LSTM ((Bi)LSTM).[19] However, only 12% of the papers used classical RNNs for FND, 2.1 Related work. This relatively limited adoption level is likely due to an infamous difficulty that is faced in RNNs - known as vanishing gradient problem, during learning of long range dependencies of textual data[].

To deal with this, researchers have used more sophisticated RNN variants like LSTM and (Bi) LSTM that incorporate memory cells and gating functions to capture information over longer sequences[34,[35].

Research focus has shifted noticeably toward these enhanced RNN variants according to our findings. More researchers are studying LSTM-based models because they excel at identifying necessary context and timebased patterns which help differentiate fake news from legitimate information [36].

Pie Chart Visualization (Suggested Breakdown):

| Model Type | Percentage |
|---|---|
| Vanilla RNN | 12% |
| LSTM | 28% |
| Bi-LSTM | 22% |
| GRU | 8% |
| Other RNN Variants | 5% |
| Total RNN-based Models | 75% |
| Non-RNN Models (e.g., CNNs, hybrids) | 25% |




Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) and has proved to work fairly well in a number of sequences related tasks like text classification and thus a good candidate for detection of fake news. In Figure 5.The GRUs have been introduced to model the temporal relationships while alleviating the vanishing gradient issue of RNNs. GRUs, on the other hand, have a smaller number of parameters (by merging the forget and input gates into an update gate) and are computationally more efficient as compared to LSTMs[37],[38].

During our proposed FND architecture with GRU implementation the textual data undergoes processing through these stages:

1) Input Layer: The textual data undergoes tokenization and padding until it reaches a standard length of 1000 tokens.
2) Embedding Layer: The model translates each word into a dense vector space of fixed size (such as 100 dimensions) to maintain semantic similarity.
3) GRU Layer: The sequence's temporal dependencies and contextual relationships are captured by using either a single GRU layer or multiple stacked GRU layers. The gates in each unit manage how information flows through update and reset functions.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue V May 2025- Available at www.ijraset.com*

*4)* Dropout Layer (Optional): A Dropout Layer is implemented post-GRU to avoid overfitting during training by randomly nullifying some activations.
*5)* Dense Layer: A fully-connected layer that takes the output of the GRU and generates a feature representation suitable for classification.
*6)* Output Layer: The last layer is a soft-max or sigmoid layer (based on either binary or multi-class setup) that gives the final prediction — whether the given news is fake or real.

This trade-off between performance and training complexity ensures that the GRU is a practical solution in FND systems.

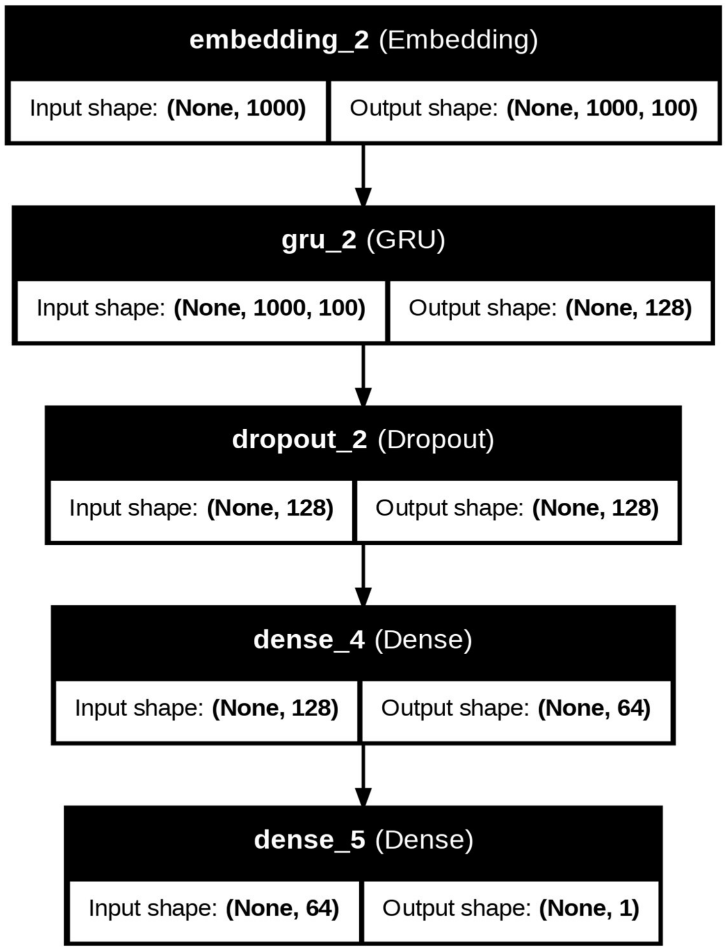



Figure 5. An example of GRU architecture used in FND.

### *C. Architectures Based On Graph Neural Networks*

Graph Neural Network (GNN) is one of the powerful methods of false news detection that can leverage the relational form of social data. In contrast to traditional neural networks that process inputs independently, GNNs process a set of nodes that are connected in a graph, making it possible for them to model the complicated relationships between users, posts and their [39] interactions.

In the context of fake news detection, a GNN is capable of modelling:

Nodes: User, News Article, Source

Edges: Actions like follows, liked, shared, re-posted, or relation of articles with their sources.

Features: Text embeddings, user reputation scores, temporal and cascading features

How GNN Helps in FND

- Captures propagation patterns of fake news across user networks
- Uses structural features (user credibility, source reputation)  Identifies community-driven anomalies in spread behavior  A typical GNN fake news detection pipeline consists of:  Graph Construction:
- Nodes = users + posts + sources

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

- Edges = user- post interactions, post-source relations  Feature Initialization:
- Post nodes: Text embeddings (TF-IDF, BERT, or LSTM)
- User nodes: Vectors of behavior, social credibility scores  Sources: Reputation score, historical records  Graph Neural Network Layers:
- Use GCN, Graph SAGE or GAT (Graph Attention Networks) layers.  Information pool from the neighbors  Pooling & Readout Layer:
- Universally Pooling Graphs We may aggregate node features into a graph representation  Classifier (MLP):
- Fully connected dense layers
- Output: Fake or Real labels on test data.

### *D. Attention-Based And Bert-Based Architectures*

One of the most powerful techniques in NLP including Fake news detection is BERT (Bidirectional Encoder Representations from Transformers). Contrary to standard models that read the text in a left-toright or a right-to-left sequence, BERT has the ability to read the full input in one go, thus being simultaneously able to capture more extensive context and relationships between words [40].

In fake news detection works, BERT-based models are mostly pre- trained and then fine-tuned on labelled datasets. The input raw text is tokenized into sub-words using BERT's Word Piece tokenizer. The tokenized input is then fed into a pre-trained BERT model to generate contextualized embeddings for each token. The representation for the special [CLS] token (that summarises the whole sequence) is sometimes used for the input of a classification layer.

This concluding layer, for example together with a softmax activation, predicts the probability that the given text is genuine or fake. BERT being pre-trained on large corpora, it can efficiently capture linguistic nuances (wit), sarcasm and implicit cues – factors which can be very useful in detecting misinformation.

Such models would be specially beneficial since they: Need little architecture engineering for task.

Achieve state-of-the-art results on several NLP tasks, Deal with Difficult Syntactic and Semantic [41].

### *E. Challenges Related To Deep Learning Methods For Fake News Detection*

Although deep learning (DL) methods have demonstrated promising results in the field of fake news detection (FND), various challenges remain that prevent their practical applications to various environments. One of the keys issues is the presence of and the quality of annotated datasets. Deep models designed to generalize to a wide range of natural image are best trained on large amounts of high-quality data, however the available FND datasets are often small, domain-specified, or imbalanced, and they do not allow the generalization capacity of the model to be efficiently exploited. Additionally, the constantly updated new fake news content cannot be fully captured by the static datasets. The problem of domain adaptation makes this an even harder problem, for models that perform well in one news sub-type (e.g., political) frequently perform poorly in another (e.g., medical to original) [43].

A second major issue with deep learning is the expensiveness (in terms of time; data; and computing resources, including memory, CPUs, and parallel processing units) of training and running deep learning models. Models like BERT, LSTM, or hybrid CNN-RNN systems are computationally intensive, require high-end GPUs and long training times, which make it hard for real-time detection and deployment in resource-constrained systems. In addition, the DL models tend to be black-box solutions, such that these networks could make a decision without being able to make inferences from it. Sensitive applications, where it is the most concerning. Where it is important to make machine learning predictions interpretable so that it is possible to hold models accountable and trust them. A DL based model also has adversarial defense and evasion limitations. Adversaries can inject malicious features that contribute to misleading the model inside humans' textual input by inserting few lexical changes or rephrasing or changing the syntax that the model is fooled by and also maintains the deceptive intention. These adversarial samples reveal the brittleness in even the most advanced models, and that there is still more work to do in terms of creating robust model architectures. The temporal and situational domains compound the problem: fake news narratives can evolve quickly over time and vary significantly by social-political environment, which are ignored by static models. As a result, non-heuristic models that are not continuously updated and contextaware will be outdated sooner due to ever-changing systems. Bias and fairness also raise important ethical and operating issues. These may accidentally perpetuate or amplify existing biases about gender, ethnicity, political ideology, or geographical origin. This bias may result in discriminative cause discriminatory [44]mis-specification and lack of faith in the model predictions.

Lastly, given that the fake news is multimodal (i.e., can combine text, image, video and links to other sources), this is a very difficult task. Most of these models are designed primarily for the text and fail to effectively bring in or reason across multiple modalities, overlooking important cues from the associated visual or multimedia information [45].

## IV. DATASETS USED FOR FAKE NEWS DETECTION

In this section, we first discuss the main characteristics of the datasets used in the surveyed works Then, we discuss some of the open challenges related to the datasets in this application domain.

### *A. Main Datasets Used For Fake News Detection*

Researchers have used several datasets in the context of fake news detection. However, we found that only a small part of these datasets is publicly available, while a considerable percentage is created by the researchers and/or is not disclosed publicly. A pie chart of the used datasets in the surveyed studies is presented in Figure 6.

### *B. LIAR Dataset*

Research on fake news frequently uses the LIAR dataset. It contains more than 12,000 brief political quotes extracted from the reliable fact-checking website PolitiFact.com. The distinctive feature of this dataset is that every statement is assigned one of six labels, ranging from "true" to "pants on fire," which is wholly untrue. It offers useful background information in addition to the text, such as the speaker's political affiliation, job title, and even their level of honesty in the past [48].

### *C. Fake and Real News Dataset*

One of the most widely used datasets on Kaggle, it is intended for binary classification, or the ability to discern between fake and authentic news. About 44,000 full-length news articles are included. Fake articles are taken from sites that are known to disseminate false information, whereas authentic ones are from reputable news organizations like Reuters [49] Models can learn patterns of dishonest versus reliable reporting because each article has a title, text content, subject, and publication date.

### *D. ISOT Fake News Dataset*

Similar to the previous dataset, this one was created by the ISOT Lab at the University of Victoria and includes more than 44,000 news articles. The fake news was collected from websites that fact-checkers had flagged, while the real news is sourced from reliable sources. Every entry includes the entire article text along with a headline. It's frequently used to test how well models can identify fake news based solely on an article's language analysis [50].

### *E. FakeNewsNet*

By fusing social media data with news content, FakeNewsNet goes one step further. It contains news stories from sites like PolitiFact and GossipCop, as well as information about how those stories were disseminated on Twitter, including user profiles, tweets, and retweets. This dataset is ideal for investigating how social behavior can be used to identify false information and how fake news propagates online. Researchers who wish to integrate text analysis with social network analysis will find it particularly helpful.

### *F. BuzzFeed News Dataset*

This dataset is the result of a BuzzFeed News project in which reporters manually verified the accuracy of Facebook posts from nine major publishers during the 2016 U.S. presidential election. Every article was classified as either mostly false, mostly true, or a combination of true and false. Despite being smaller than other datasets, this one is especially helpful for studying political news and disinformation on social media, particularly during election seasons.

### *G. COVID-19 Fake News Dataset*

This dataset, which was produced at the height of the COVID-19 pandemic, focuses exclusively on false information about health. It contains a range of social media posts and news articles that are classified as either authentic or fraudulent. This dataset is frequently used for studies that seek to identify false claims regarding medical treatments, vaccines, and the virus itself because of how crucial accurate information is during a health emergency [51].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

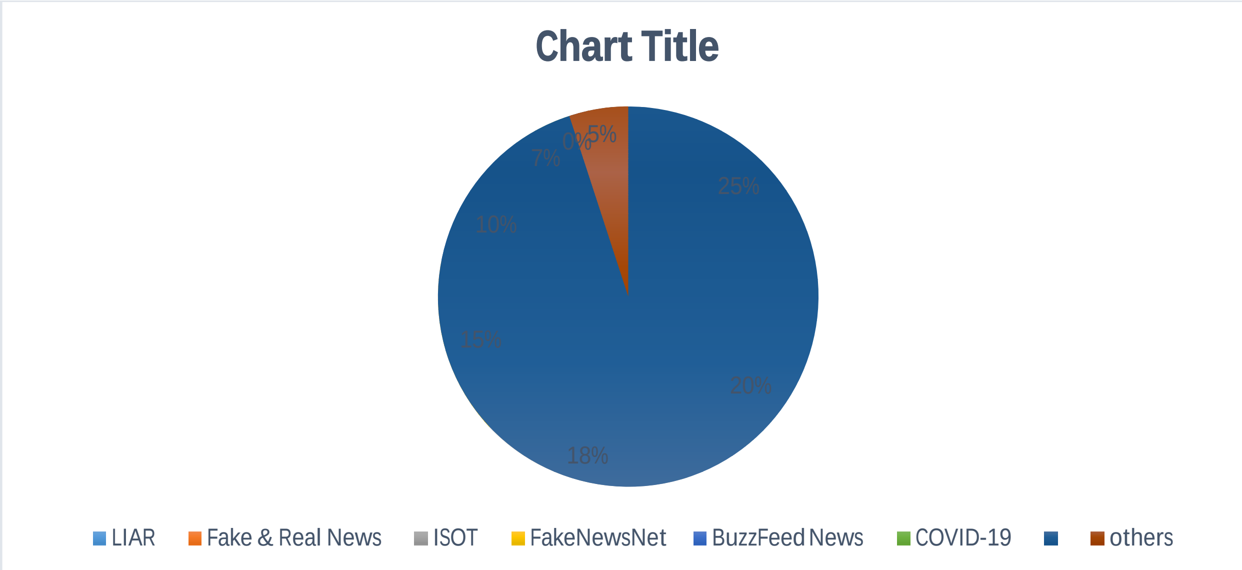



Figure 6. Datasets used in the surveyed studies.

| Dataset | Size | Content Type | Labels | Domain | Additional Features |
|---|---|---|---|---|---|
| LIAR | 12,836 statements | Political claims | 6-class (True to Pants-fire) | Politics | Speaker metadata, context, credibility history |
| Fake & Real News | ~44,000 articles | Full news articles | Binary (Fake, Real) | General news | Article titles, subjects, and text |
| ISOT | ~44,898 articles | Full news articles | Binary (Fake, Real) | Politics, World News | Title + full content, clean structure |
| FakeNewsNet | Thousands of articles | News + social media posts | Binary (Fake, Real) | Politics, Gossip | Twitter interactions, user profiles, propagation data |
| BuzzFeed News | ~2,000 articles | Facebook articles | 3-class (Mostly True, Mixed, False) | Politics | Manually annotated by journalists during the 2016 US election |
| *(Unnamed – COVID)* | ~10k–20k | News + social media | Binary (Fake, Real) | Health | Health misinformation curated from WHO-verified sources |

Table 1. Main characteristics of the publicly available FND datasets.

## V. ANSWERS TO RESEARCH QUESTIONS

*1) RQ1: What is fake news and how it does affect people and society?*

The term "fake news" describes inaccurate or misleading information that is passed off as real news, frequently with the intention of misleading or controlling. It has an impact on people by disseminating false information, forming skewed viewpoints, and undermining media credibility. It can affect elections, exacerbate polarization, and jeopardize democratic institutions and public health on a societal level.

*2) RQ2: Which deep learning algorithms have been used for fake news detection throughout time?*

Several deep learning algorithms have been used to detect fake news over time. For text classification, early models employed CNNs and RNNs, particularly LSTM and GRU. Bi-LSTM and attention mechanisms later improved performance. Transformer-based models, such as BERT, RoBERTa, and XLNet, have recently taken the lead because of their exceptional accuracy and contextual awareness. For better detection, these models are frequently refined on datasets of fake news.

*3) RQ3: How effective are deep learning methods for fake news detection?*

Deep learning methods have proven highly effective for fake news detection due to their ability to capture complex patterns in language and context. Models like LSTM, CNN, and transformers (e.g., BERT) achieve high accuracy and outperform traditional machine learning approaches. Their effectiveness improves further when combined with transfer learning and domain-specific fine-tuning.

*4) RQ4: How to create a pertinent model for detecting fake news?*

There are several crucial steps involved in developing a relevant model for fake news detection. Choose a high-quality labeled dataset first (LIAR, ISOT, etc.). Next, perform preprocessing on the text (tokenization, cleaning). Select a deep learning model that works well; transformer-based models, such as BERT, are frequently employed. To improve performance, fine-tune the model using your dataset and apply strategies like transfer learning. Measures like accuracy, precision, recall, and F1-score are used for evaluation. Model relevance and dependability can be further increased by incorporating metadata (such as user behavior and source credibility) and making sure that the domain is adapted.

*5) RQ5: How to prevent the spread of fake news?*

A mix of technological, educational, and regulatory strategies are needed to stop the spread of fake news. In terms of technology, social media platforms can flag or block deceptive content by implementing precise fake news detection systems. Users who receive media literacy training are better able to assess information sources critically. Claims can be verified with the help of browser extensions and fact-checking services. Promoting trustworthy journalism and fostering responsible sharing practices are also important. Governments and platforms can also enact laws and rules to punish willful disinformation and guarantee openness in content moderation.

## VI. MAIN GAPS AND OPEN ISSUES

Even with great advancements, there are still a number of holes and unresolved problems with fake news detection. The absence of high-quality, varied datasets from various fields, languages, and cultures is a significant obstacle that restricts the generalizability of models. Multimodal misinformation is more difficult to detect because most detection systems concentrate on textual content, but fake news nowadays frequently contains images, videos, and memes. The inability of current models to detect fake news before it becomes widely disseminated raises additional concerns regarding real-time detection [52], [53],[54],[55].

## VII. CONCLUSION

One important and developing field in social media and artificial intelligence research is the identification of fake news. Public opinion, democratic processes, and societal trust are all significantly impacted by the spread of false information. Fake content has been successfully identified by deep learning models, particularly transformer-based architectures like BERT and RoBERTa. Performance has been further improved by transfer learning, especially in situations where labeled data is scarce. Significant obstacles still exist, though, including the requirement for real-time detection systems, the lack of diverse and multilingual datasets, and the restricted support for multimodal content (text, images, and videos). Furthermore, trust and explainability are questioned due to deep models' black-box nature. It's equally critical to address moral dilemmas like bias, censorship, and user privacy. To develop ethical, scalable, and reliable fake news detection.

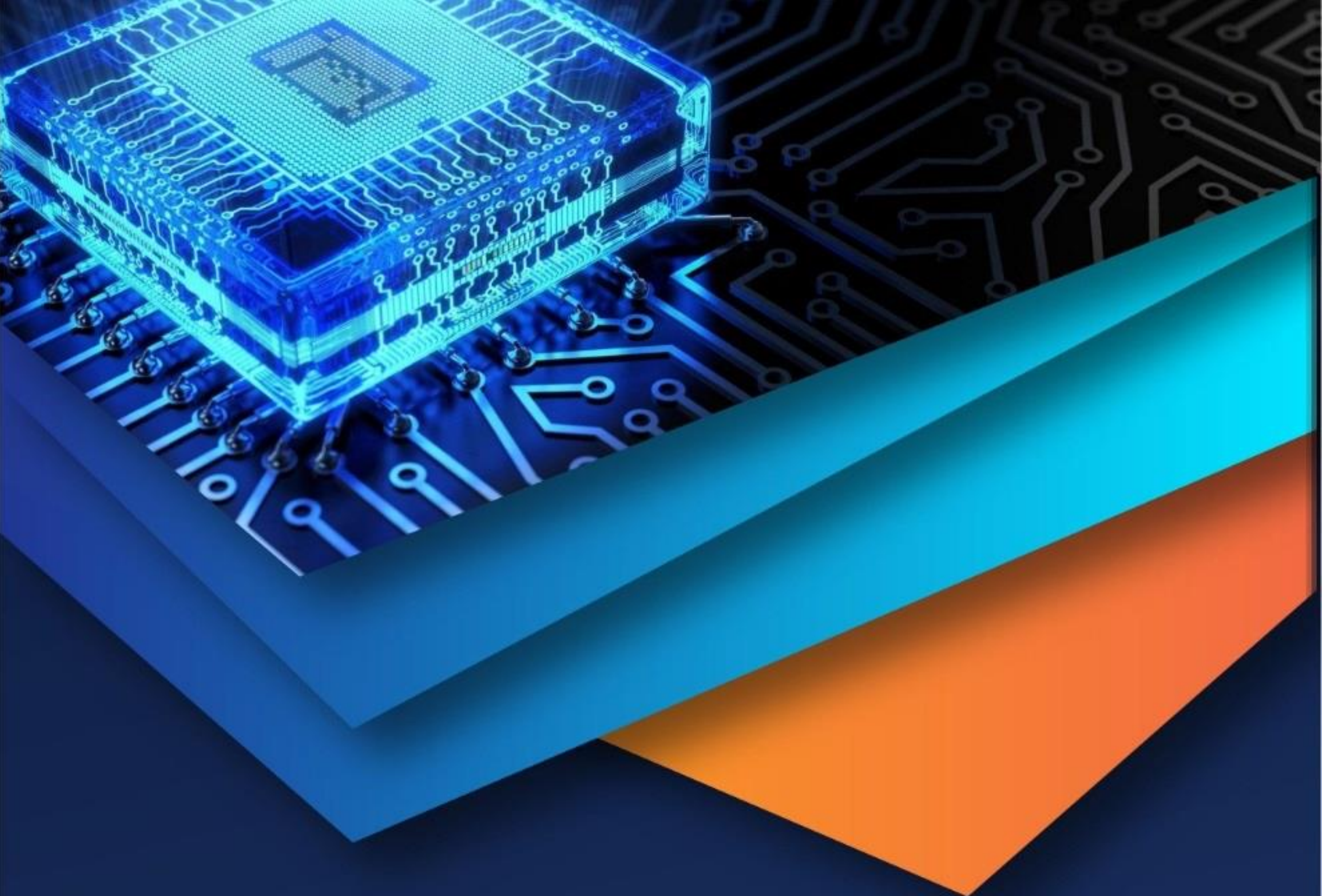
But significant obstacles still exist in spite of these developments. These include the growing need for multimodal fake news detection, the lack of diverse, multilingual, and real-world datasets, and the need for real-time response systems. Deep learning models' black-box nature raises questions about interpretability, accountability, and fairness—all of which are crucial for maintaining public confidence and moral application. Furthermore, models need to be flexible enough to adjust to new platforms and patterns as disinformation strategies change constantly.

The creation of transparent, explicable AI systems, strong multimodal integration, and cross-domain and cross-lingual generalization must be the top priorities of future research. To create effective countermeasures, researchers, legislators, and tech companies must work together. To stop the spread of fake news, it is equally important to educate the public, improve media literacy, and encourage responsible digital behavior. In the end, developing reliable, scalable, and socially conscious fake news detection systems will require an all-encompassing, multidisciplinary approach.

## REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017.
[2] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," arXiv preprint arXiv:1812.00315, 2018.
[3] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," Commun. ACM, vol. 59, no. 7, pp. 96–104, 2016.
[4] G. Pennycook and D. G. Rand, "The psychology of fake news," Trends Cogn. Sci., vol. 25, no. 5, pp. 388–402, 2021.
[5] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Secur. Privacy, vol. 1, no. 1, p. e9, 2018.
[6] R. Mihalcea and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in Proc. ACL-IJCNLP, 2009, pp. 309–312.
[7] X. Zhou, A. Jain, and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.
[8] B. D. Horne and S. Adalı, "This just in: Fake news packs a lot in title, uses simpler, repetitive content," in Proc. ICWSM, 2017.
[9] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, 2017.
[10] E. Ferrara, "What types of COVID-19 conspiracies are populated by Twitter bots?," First Monday, vol. 25, no. 6, 2020.
[11] C.-O. Truică and E.-S. Apostol, "MisRoBÆRTa: Transformers versus Misinformation," arXiv preprint arXiv:2302.00001, 2023.
[12] H. Wang et al., "Attacking Fake News Detectors via Manipulating News Social Engagement," arXiv preprint arXiv:2304.00002, 2023.
[13] C.-O. Truică, E.-S. Apostol, and P. Karras, "DANES: Deep Neural Network Ensemble Architecture," arXiv preprint arXiv:2306.00003, 2023.
[14] A. D. Radhi et al., "A comprehensive review of machine learning-based models for fake news detection," BIO Web Conf., vol. 97, p. 00123, 2024.
[15] S. G. Arkaan, A. R. Atmadja, and M. D. Firdaus, "Fake News Detection in the 2024 Indonesian General Election Using BI-LSTM," Komputasi, 2024.
[16] A. Bhardwaj and S. Kim, "Fake social media news and distorted campaign detection framework using sentiment analysis," Heliyon, vol. 10, no. 16, p. e36049, 2024.
[17] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Fake News Detection and Classification," Future Internet, vol. 17, no. 1, p. 28, 2025.
[18] S. Harris et al., "Fake News Detection Revisited," Technologies, vol. 12, no. 11, p. 222, 2024.
[19] A. A. Olagunju and I. O. Awoyelu, "Performance Evaluation of Fake News Detection Models," Int. J. Inf. Technol. Comput. Sci., vol. 16, no. 6, pp. 89–100, 2024.
[20] Cyabra, "Cyabra: AI-powered platform combating disinformation," 2024. [Online]. Available:
[21] https://www.cyabra.com
[22] YouScan, "YouScan: Social media monitoring and analytics platform," 2025. [Online]. Available:
[23] https://www.youscan.io
[24] Logically, "Logically: AI-powered fact-checking and misinformation detection," 2025. [Online]. Available: https://www.logically.ai
[25] Graphika, "Graphika: Social network analysis company tracking online disinformation," 2025. [Online]. Available: https://www.graphika.com
[26] T. S. Camelia, F. R. Fahim, and M. M. Anwar, "A Regularized LSTM Method for Detecting Fake News Articles," arXiv, 2024.
[27] H. Chen et al., "A Self-Learning Multimodal Approach for Fake News Detection," arXiv, 2024. [26] J. Singh et al., "LingML: Linguistic-Informed Machine Learning for Enhanced Fake News Detection," arXiv, 2024.
[28] J. Su, C. Cardie, and P. Nakov, "Adapting Fake News Detection to the Era of Large Language Models," arXiv, 2023.
[29] J. Pastor-Galindo and J. A. Ruipérez-Valiente, "LLM-Powered Agent-Based Framework for Misinformation Research," arXiv, 2023.
[30] J. Tong and J. Marx, "Combatting Deepfake Misinformation on Social Media: A Scoping Review and Research Agenda," ResearchGate, 2024.
[31] S. Bose, "Exploding AI-Generated Deepfakes and Misinformation: A Threat to Global Security," SSRN, 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4651093
[32] S. Patil, "Misinformation Detection on Social Media using AI and Machine Learning," Int. J. Sci. Res. (IJSR), vol. 13, no. 5, pp. 1013–1015, 2024.
[33] A. Gangurde et al., "Convergence of Artificial Intelligence with Social Media: A Comprehensive Review," ScienceDirect, 2024.
[34] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," Complexity, 2020.
[35] X. Zhou, R. Zafarani, K. Shu, and H. Liu, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Comput. Surv., vol. 52, no. 3, pp. 1–40, 2019.
[36] R. Oshikawa, J. Benjamin, and E. Choi, "A Survey on Natural Language Processing for Fake News Detection," arXiv preprint arXiv:2007.00846, 2020.
[37] S. Yuan, J. Zhang, and L. Yao, "Fake News Detection on Social Networks: A Survey," Expert Syst. Appl., vol. 213, p. 118773, 2023.
[38] D. M. Villela, F. A. de Oliveira, and R. R. Ferreira, "Fake News Detection: A Systematic Literature Review of Machine Learning Algorithms and Datasets," Appl. Soft Comput., vol. 132, p. 109869, 2023.
[39] Anonymous, "Fake News Detection: It's All in the Data!," arXiv preprint, 2024.
[40] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," arXiv preprint arXiv:1705.00648, 2017.
[41] K. Shu, S. Wang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017.
[42] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FakeBERT: Fake News Detection in Social Media with a BERT-based Deep Learning Approach," J. Inf. Process. Syst., vol. 17, no. 2, 2021. [42] K. Shu, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," Big Data, vol. 8, no. 3, pp. 171–188, 2020.
[43] Anonymous, "Detecting Fake News using Machine Learning: A Systematic Literature Review," arXiv preprint, 2021.
[44] S. Balshetwar and S. Sawant, "Fake news detection in social media based on sentiment analysis using classifier techniques," Int. J. Electr. Comput. Eng. (IJECE), vol. 13, no. 3, pp. 3126–3133, 2023. [45] M. M. Hamed, A. Darwish, and A. E. Hassanien, "A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion," Artif. Intell. Rev., vol. 56, no. 1, pp. 1–53, 2023.

[45] Anonymous, "Deep Learning Based Fake News Detection on Social Media," DergiPark, 2023. [47] G. Al-Tarawneh, M. Al-Smadi, M. A. Jarrah, and A. A. Shatnawi, "Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach," Int. J. Intell. Syst. Appl. Eng., vol. 11, no. 3s, pp. 355–366, 2023.
[46] A. D'Ulizia, E. Ferrara, and A. Fazzini, "Fake news detection: a survey of evaluation datasets," J. Big Data, vol. 8, no. 1, pp. 1–20, 2021.
[47] V. Shah and S. Patel, "A Comprehensive Survey on Fake News Detection Using Machine Learning," 2025.
[48] S. Yuan, J. Zhang, and L. Yao, "Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice," Environ. Sci. Pollut. Res., vol. 30, no. 10, pp. 26687–26707, 2023.
[49] A. Jain, "A smart System for Fake News Detection Using Machine Learning," ResearchGate Preprint, 2020.
[50] Anonymous, "Fake News Detection: A Survey," IRJMETS, 2021.
[51] Anonymous, "Fake News Detection: A Comprehensive Study on Analysis and Modern Classification Techniques," ResearchGate, 2024.
[52] H. Karimi and J. Tang, "Text-mining-based Fake News Detection Using Ensemble Methods," Inf. Process. Manag., vol. 55, no. 6, pp. 2043–2059, 2019.
[53] G. K. Shahi, D. Nandini, and R. Chandrasekaran, "Covid-19 Fake News Detection: A Survey," Appl. Intell., vol. 51, no. 10, pp. 7277–7305, 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY