**Engineering and Technology Journal e-ISSN: 2456-3358**

**Volume 10 Issue 05 May-2025, Page No.- 5087-5099**

**DOI: 10.47191/etj/v10i05.37, I.F. – 8.482**



# The Role of NLP In Fake News Detection and Misinformation Mitigation

**Shahwan younis ali[1], Dr.Ibrahim Mahmood[2]**

[1]Information Technology, Technical College of Informatics, Akre University for Applied Science, Kurdistan Region, Iraq

[2]Information Technology Management, Technical College of Informatics, Akre University for Applied Science, Kurdistan Region, Iraq

**ABSTRACT:** In the era of rapid digital communication, the spread of fake news has emerged as a global challenge, influencing public opinion and undermining trust in information systems. This review explores the pivotal role of Natural Language Processing (NLP) in detecting fake news and mitigating misinformation. Various approaches integrate NLP techniques with machine learning (ML) and deep learning (DL) architectures to enhance detection accuracy and robustness. Commonly used text representation methods include TF-IDF, Word2Vec, GloVe, and BERT, often supported by syntactic and semantic features such as POS tagging, named entity recognition (NER), stylometry, and sentiment analysis. Advanced architectures like CNN-RNN hybrids, dual BERT models, and capsule networks have demonstrated high effectiveness, with performance metrics reaching up to 99.8% on benchmark datasets. Further strategies such as ensemble learning, stance detection, adversarial robustness, and the incorporation of external verification tools have been shown to improve credibility assessment. While many models achieve high accuracy in controlled environments, challenges persist in cross-domain generalization, multilingual adaptability, and ethical transparency. This review highlights the critical contributions of NLP in combating misinformation and recommends future systems to leverage multimodal data, real-time responsiveness, and explainable AI for more resilient and trustworthy detection frameworks.

**KEYWORDS:** Fake News Detection, Natural Language Processing, Deep Learning, Stylometry, Misinformation Mitigation.

## 1. INTRODUCTION

The digital era has witnessed an unprecedented surge in the dissemination of news via social media and online platforms, inadvertently creating fertile ground for the spread of fake news and misinformation. Natural Language Processing (NLP), a subfield of artificial intelligence focused on enabling machines to understand human language, has emerged as a critical tool in combating this growing challenge [1]. Fake news often mimics legitimate journalistic content, making manual verification difficult and time-consuming. NLP offers scalable, automated approaches to evaluate content credibility, analyze linguistic patterns, and identify deceptive or manipulated narratives[2][3].

Several studies have highlighted the integration of NLP with deep learning techniques such as LSTM, BERT embeddings, and attention mechanisms to enhance fake news detection accuracy [4][5]. These models exploit the semantic and contextual characteristics of text, enabling the differentiation between authentic and fabricated information. Researchers like [6] emphasized the efficacy of hybrid NLP-based frameworks in processing large volumes of data and identifying deceptive patterns across diverse domains.

Misinformation, particularly when spread systematically, poses significant threats to democratic institutions, public health, and social cohesion. The inability of users to discern truth from falsehood, often due to cognitive biases and algorithmic echo chambers, amplifies the impact of such content [7] . NLP-driven fake news detection systems not only aim to classify news as real or fake but also assist in tracking sources, verifying facts, and contextualizing content within verified information structures [8][9] .

Thus, the role of NLP in fake news detection is not merely technical—it is foundational in restoring information integrity, enhancing media literacy, and supporting fact-based discourse in digital spaces

## 2. RESEARCH METHODOLOGY

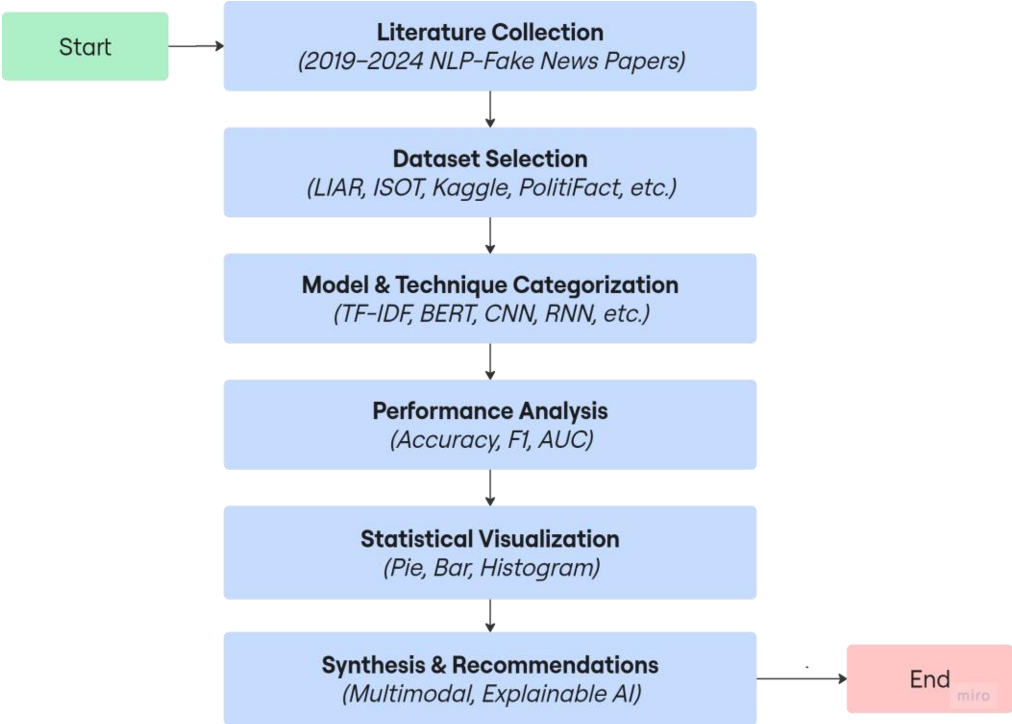


**Figure 1: Flowchart: Research Methodology for NLP in Fake News Detection**

This study adopts a systematic review-based methodology to explore the integration and effectiveness of Natural Language Processing (NLP) techniques in fake news detection and misinformation mitigation. The research was conducted in three stages: dataset selection, literature synthesis, and comparative analysis.

### 2.1 Data Collection and Literature Selection

Relevant peer-reviewed articles published between 2019 and 2024 were collected, focusing on fake news detection systems employing NLP and machine learning (ML) or deep learning (DL) techniques. The selection criteria included studies with publicly available datasets such as LIAR, ISOT, FakeNewsNet, Kaggle, and PolitiFact, which provided real-world news and social media content for model training and evaluation.

### 2.2 Model and Technique Categorization

The reviewed studies were categorized based on their core NLP techniques and model architectures. These included:

- **Text Representation Techniques**: TF-IDF, Word2Vec, GloVe, and BERT embeddings.
- **Classification Models**: Logistic Regression, SVM, Random Forest, Naive Bayes, CNN, RNN, LSTM, BiLSTM, Capsule Networks, and Transformer-based models (e.g., BERT, GPT-3.5).
- **Hybrid and Ensemble Strategies**: CNN-RNN hybrids, BERT with attention mechanisms, Soft Voting, and Boosting.

Each study's approach to preprocessing—such as tokenization, POS tagging, lemmatization, and stylometric analysis—was noted, and the use of external tools like NER, sentiment analysis APIs, and fact-checking resources was documented.

### 2.3 Performance Comparison and Statistical Analysis

Studies were compared using key performance metrics: accuracy, precision, recall, F1-score, AUC, and adversarial robustness. Graphical analyses, including bar charts, pie charts, and histograms, were used to visualize:

- The distribution of research objectives (e.g., hybrid model development, stylometric profiling, stance detection),
- The most frequently used NLP techniques,
- The accuracy performance range of different models.

High-performing models like FNDNet (98.36% accuracy), BERT-based systems (up to 98%), and CNN-RNN hybrids (up to 99%) were benchmarked against traditional ML classifiers to highlight advancements in NLP-based detection systems.

### 2.4 Synthesis and Recommendation Framework

Based on observed trends, the methodology extended to formulating evidence-based recommendations. These were centered on integrating multimodal content, enhancing model generalizability across domains and languages, adopting explainable AI, and addressing ethical considerations in model deployment.

## 3. BACKGROUND THEORY

### 3.1. Introduction to Fake News and Misinformation

In the digital age, the proliferation of content across online platforms has drastically changed how individuals consume information. Amid this shift, **fake news** has emerged as a pervasive challenge, defined as fabricated information that mimics the style and format of legitimate journalism but lacks factual accuracy and is often spread with deceptive intent [10] . Closely related are **misinformation**, which refers to unintentionally false or misleading information, and **disinformation**, which involves the deliberate dissemination of false content to manipulate public opinion or behavior [11] .

The impacts of fake news span multiple domains. **Socially**, it exacerbates polarization by reinforcing pre-existing biases and creating echo chambers. **Politically**, fake news has been implicated in undermining democratic processes, as seen in events like the 2016 U.S. elections and various misinformation campaigns targeting voters. **Economically**, market behaviors have been manipulated by viral fake headlines, such as false reports of corporate bankruptcies or geopolitical crises that influence stock market volatility [3]. In the realm of **public health**, misinformation during the COVID-19 pandemic severely hindered vaccination campaigns and public compliance with safety protocols [12].

A critical challenge is the **speed and scale** at which fake news spreads, particularly through **social media** platforms like Facebook, Twitter, and YouTube. Algorithms that prioritize user engagement often amplify sensational content, making fake news more likely to go viral than accurate reporting [13]. Additionally, the lack of centralized editorial control and real-time fact-checking mechanisms on these platforms enables widespread dissemination of unverified information. The presence of **social bots** and coordinated disinformation networks further compounds the problem, facilitating the artificial boosting of misleading narratives [14].

### 3.2. The Role of Natural Language Processing (NLP)

**Natural Language Processing (NLP)** is a subfield of artificial intelligence and computational linguistics concerned with enabling machines to process and understand human language. It provides a suite of techniques for analyzing textual data, making it invaluable for identifying semantic and syntactic patterns in fake news articles.

The core advantage of NLP in this context is its ability to **analyze content directly**, independent of metadata or source credibility. This enables the identification of subtle cues such as exaggeration, emotional manipulation, or biased language. NLP techniques help uncover **linguistic markers** of deception, such as excessive adverbs, passive voice, or emotionally charged phrases, which often characterize fabricated stories [15].

Key NLP tasks applicable to fake news detection include:

- **Text Classification** – categorizing news as real or fake.
- **Sentiment Analysis** – assessing emotional tone, which can indicate manipulative intent.
- **Named Entity Recognition (NER)** – extracting and evaluating people, places, and organizations mentioned in the article.
- **Part-of-Speech (POS) Tagging** – identifying grammatical structure to understand writing style.
- **Semantic and Syntactic Parsing** – analyzing sentence structure and meaning.

These tasks form the foundational building blocks for automated fake news detection systems and enable models to make fine-grained distinctions between authentic and deceptive content.

### 3.3. NLP Techniques Used in Fake News Detection

The performance of NLP systems largely depends on how textual information is represented and modeled. Early models relied on **term frequency-inverse document frequency (TF-IDF)** and **bag-of-words** representations to extract features. Later, **word embedding techniques** such as **Word2Vec**, **GloVe**, and more recently, **contextualized embeddings** like **BERT** and **GPT** provided deeper semantic understanding by preserving word context [16][17].

Advanced models also incorporate **stylometric analysis**, which focuses on writing styles and linguistic patterns such as punctuation usage, lexical richness, and sentence complexity. These features are particularly useful for **author profiling** and identifying content that deviates from normative journalistic standards.

Additionally, **transformer-based models** such as BERT, ELMo, and RoBERTa use attention mechanisms to focus on relevant parts of the input, enabling deeper contextual interpretation. These models significantly outperform earlier architectures in fake news classification tasks [4].

Other important approaches include:

- **Behavioral Profiling**: Examining patterns in content sharing, such as time of posting or network diffusion.
- **Stylometric Profiling**: Using features like pronoun frequency, emoji usage, or punctuation density to model the writing persona.
- **Multimodal Fusion**: Combining textual, visual, and user behavioral data for holistic analysis.

### 3.4. Machine Learning and Deep Learning Models in NLP

Machine learning and deep learning serve as the backbone of most NLP-powered fake news detectors. Traditional algorithms like **Logistic Regression**, **Naive Bayes**, and **Support Vector Machines (SVM)** are still used for lightweight classification tasks but often struggle with capturing nuanced textual dependencies.

Modern systems increasingly rely on **deep learning architectures**, such as:

- **Convolutional Neural Networks (CNNs)**: Useful for extracting spatial hierarchies in text.

- **Recurrent Neural Networks (RNNs)**: Suitable for sequence modeling, especially when paired with **LSTMs** or **GRUs** for long-term dependency capture.
- **Bi-directional LSTM (BiLSTM)**: Allows contextual information to flow in both directions across the text, enhancing semantic accuracy.
- **Capsule Networks and Attention Mechanisms**: Improve the learning of positional and relational information in text.

**Hybrid models**, such as **CNN-BiLSTM with attention layers**, have shown superior performance in handling fake news datasets. These models excel at extracting both syntactic features (via CNNs) and temporal or contextual features (via LSTMs), with attention modules enabling models to focus on linguistically significant phrases [18][19].

### 3.5. NLP-Driven Fake News Verification Systems

Beyond classification, NLP is also used for **automated verification** of claims. Integration with external fact-checking tools (e.g., **ClaimBuster**, **Google Fact Check Tools**) allows systems to cross-reference content with verified knowledge bases. **Knowledge graphs**, **semantic reasoning engines**, and **retrieval-based verification models** are commonly embedded into NLP pipelines to validate statements and expose logical inconsistencies.

Some systems leverage **discourse parsing** and **entity co-reference resolution** to identify contradictions or false attributions within the narrative itself [11]. Others employ **retrieval-based QA** models to fetch relevant facts and match them against article content for consistency checking.

## 4. LITERATURE REVIEW

[14] used the "ISOT Fake News Dataset" from the University of Victoria to detect fake news through a detailed process that involved text preprocessing, the application of NLP techniques like stemming and TF-IDF, and machine learning. The study employed various classifiers and ensemble methods, testing them on a split dataset to evaluate their accuracy, precision, recall, and F1-score. The effectiveness of these models was compared by analyzing both news titles and full texts to see if simpler methods matched the accuracy of more complex approaches, showcasing an efficient integration of NLP and machine learning for fake news detection.

[20]utilized the SEMEVAL-2019 by-article dataset containing 1273 news articles and applied data preprocessing to remove HTML tags and unnecessary characters while retaining punctuation to preserve the style. It employed three NLP models—BERT, ELMo, and Word2Vec—for hyperpartisan news detection. BERT used an 'uncased' tokenizer for text tokenization and managed sentence lengths with padding, while ELMo produced word embeddings using a bi-directional LSTM, and Word2Vec generated embeddings from a pre-trained model, filtering out noise such as stop words. Each model was trained on 80% of the data and validated on the remaining 20%, adjusting for class imbalances with a Random Forest classifier. The study evaluated model performance through precision, recall, and F1-score to fine-tune hyperpartisan news detection.

[21]proposed deep learning models to detect fake news using datasets collected from Twitter and PolitiFact. After preprocessing and embedding the data with GloVe, they tested various models and found that a CNN combined with Bidirectional LSTM and an attention mechanism achieved the highest accuracy of 88.78%. This outperformed traditional machine learning methods, and statistical analysis confirmed the significance. The study concluded that deep learning was effective for fake news detection, with future improvements possible through the integration of source credibility.

[22] proposed a hybrid machine learning framework for fake news detection that combined NLP features with knowledge verification techniques. They introduced a flexible definition of fake news based on factual accuracy and source reliability, and highlighted the limitations of existing detection methods. Their model used linguistic features, sentiment analysis, and verification from trusted sources to assess news credibility. Tested on the Kaggle Fake News Dataset, the system showed high accuracy in distinguishing real from fake news. The authors concluded that combining automated and manual verification improved detection and recommended further exploration of semantic similarity methods.

[23] proposed a hybrid fake news detection framework that combined machine learning with semantic features such as sentiment, named entities, and relational data from knowledge graphs. Using tools like TextBlob and DBpedia, they enriched text inputs and applied models like GRU with attention and Capsule Networks on the LIAR and PolitiFact datasets. Their approach showed improved accuracy, with key words like “percent” and “president” influencing predictions. The study concluded that semantic features significantly enhanced detection performance and suggested future research using graph neural networks and expanded datasets.

[18] proposed capsule neural networks for fake news detection, using static and non-static word embeddings with n-gram feature extraction. They tested their models on the ISOT and LIAR datasets, where the models outperformed traditional methods, achieving 99.8% accuracy on ISOT and up to 40.9% on LIAR. The approach proved effective, especially when enriched with metadata, and showed promise for future enhancements using semantic and graph-based methods.

[24] proposed a CNN-LSTM model for fake news stance detection using the FNC-1 dataset. They applied preprocessing and dimensionality reduction techniques, including PCA and Chi-square, to improve performance. The PCA-based model achieved the highest accuracy of 97.8%, outperforming models like BERT, XLNet, and RoBERTa.

The approach proved effective in handling imbalanced data and enhancing classification across stance categories. The authors suggested future work that involved multilingual data and advanced learning techniques.

[25] proposed MWPBert, a dual BERT model for fake news detection that encoded both headlines and fact-check-worthy spans of news bodies using the MaxWorth algorithm. The model outperformed existing methods on the FakeNewsNet dataset, achieving 85.4% accuracy and a 0.64 F1-score. It demonstrated that parallel BERTs, MaxWorth-based span selection, and simple output layers improved performance. The study suggested future work on multilingual and multimodal extensions.

[26]proposed FactAgent, an agentic framework that used GPT-3.5 for fake news detection through a structured workflow that mimicked human fact-checking. It combined internal LLM knowledge with external tools such as SerpAPI and URL checkers. Tested on Snopes, PolitiFact, and GossipCop, FactAgent outperformed models like BERT and LSTM, achieving F1-scores of up to 0.88. It also provided transparent reasoning and demonstrated that expert-designed workflows and checklist-based decisions improved accuracy. The study suggested future enhancements, including full article and multimodal analysis.

[27] proposed a two-phase fake news detection framework that combined deep learning and NLP techniques. In the first phase, they used GloVe and BERT embeddings with MLP, CNN, and BiLSTM models, while in the second phase, they applied ensemble methods to improve accuracy. They tested the framework on four datasets, where the Soft Weighted Voting Ensemble achieved the best results on three, with F1-scores reaching up to 95.22%. The study demonstrated that ensemble learning enhanced detection performance, especially for longer texts, and suggested future work on incorporating non-textual and multilingual features.

[28]proposed a fake news detection system using a hybrid CNN-BILSTM model with combined word embeddings, including Word2Vec (CBOW and Skip-Gram) and GloVe. They trained the model on the WELFake dataset and achieved a top accuracy of 97.74% using the Word2Vec combination, which outperformed traditional machine learning methods. The study showed that combining embeddings improved classification performance and suggested future enhancements through dimensionality reduction and the integration of non-textual features.

[29] proposed a fake news detection model using a hybrid CNN-LSTM architecture trained on the Kaggle dataset. The model utilized GloVe word embeddings and processed both titles and article bodies to classify news as fake or real. With optimized parameters, it achieved 94.71% test accuracy and 97.21% precision. The CNN captured textual features while the LSTM handled sequential dependencies. The inclusion of dropout reduced overfitting. The study concluded that their model outperformed traditional methods and suggested incorporating metadata and real-time credibility tools in future work.

[30] evaluated the robustness of fake news detectors against black-box adversarial attacks using deep learning models (MLP, CNN, RNN, Hybrid CNN-RNN) on the Kaggle, ISOT, and LIAR datasets. Using attacks like Text-Bugger and Text-Fooler, they found that RNNs and CNNs were more robust than MLP and Hybrid models. Longer input sequences and binary cross-entropy loss improved robustness, though at the cost of accuracy. The study also revealed vulnerabilities to dataset bias and poor generalization across regions. It recommended developing adaptive, explainable, and cross-domain detectors in the future.

[31]proposed a fake news detection approach using language- and platform-independent text features. They evaluated Bag-of-Words, Word2Vec, DCDistance, and custom features across five multilingual datasets using classifiers like SVM and Random Forest. Bag-of-Words achieved the best performance, while DCDistance offered efficient dimensionality reduction with competitive results. The study confirmed that language-agnostic features were effective and recommended future work involving more languages, metadata, and real-time detection.

[32] developed a fake news detection system using the LIAR dataset and various machine learning algorithms, including XGBoost, SVM, and Random Forest. The system combined NLP techniques such as POS tagging and TF-IDF feature extraction. XGBoost achieved the highest accuracy at over 75%, followed by SVM and Random Forest. The study highlighted the effectiveness of combining NLP with multiple classifiers and suggested future improvements using additional POS-based statistical features.

[33] developed a fake news detection system using machine learning, deep learning, and NLP techniques on a Kaggle dataset of 26,000 articles. After preprocessing the data, they applied models including logistic regression, naive Bayes, SVM, decision tree, LSTM, and BERT. BERT achieved the highest accuracy of 98%, followed by LSTM at 95%, and outperformed all traditional ML models. The study concluded that NLP combined with deep learning, particularly BERT, was highly effective for fake news detection, and suggested future work involving topic-specific and multimodal data.

[34] proposed the NER-SA model, which combined NLP and named entity recognition for stylometric fake news detection. The model analyzed entity patterns using mathematical metrics and simulated annealing to classify news as real or fake. Tested on multiple datasets, it outperformed traditional ML and DL models, showing up to 50.18% accuracy improvement in cross-domain settings. While effective, the model faced challenges like overlapping entity use and lacked ethical considerations. The study recommended future improvements including multilingual support, real-time scalability, and enhanced adversarial robustness.

[35] proposed FNDNet, a deep CNN-based model for fake news detection that used GloVe embeddings and avoided

manual feature engineering. Tested on the Kaggle fake news dataset, FNDNet achieved a 98.36% accuracy, outperforming both traditional ML models and deep learning baselines like LSTM and CNN. The model also recorded a high true negative rate (99.41%) and a low false positive rate (0.59%). The authors suggested future improvements using transformer models like BERT and multi-modal inputs for enhanced detection.

[19] proposed a hybrid CNN-RNN model for fake news detection that combined local and sequential text feature extraction using GloVe embeddings. The model achieved 99% accuracy on the ISOT dataset, outperforming traditional ML models, and 60% accuracy on the smaller FA-KES dataset. While it was effective within datasets, it showed poor generalization across domains. The study highlighted the model's potential and recommended future improvements using dropout, bio-inspired optimization, and multimodal data integration.

[36] developed a system for profiling fake news spreaders on Twitter using stylometric and lexical features in English and Spanish. They used machine learning classifiers, achieving 72.5% accuracy with Logistic Regression for Spanish and 59.5% with Random Forest for English. Spanish tweets showed stronger stylistic patterns, which improved model performance. The study demonstrated that combining stylometric and lexical cues was effective for multilingual fake news spreader detection.

## 5. DISCUSSION AND COMPARISON

**Table 1: Comparison among the reviewed works.**

| Authors and Year | Objective | NLP Techniques Used | | Data Sources | Performance Metrics | Key Findings | Challenges and Limitations |
|---|---|---|---|---|---|---|---|
| Barbara Probierza, Piotr Stefański, Jan Kozaka, 2021 | Develop a rapid fake news detection method using ML on news titles. | Tokenization, TF, TF-IDF, stemming, lemmatization. | | ISOT Fake News Dataset, University of Victoria, Canada. | Accuracy, precision, recall, F1-score | Effective classification using only news titles. | Challenges include linguistic variability and model overfitting. |
| Navakanth Reddy Naredla, Festus Fatai Adedoyin, 2022 | Develop a model to detect hyperpartisan news articles. | BERT, ELMo, Word2Vec | | SEMEVAL-2019 by-article dataset with 1273 news articles. | Precision, Recall, F1-Score | ELMo with Random Forest had the highest accuracy (0.88); BERT and Word2Vec had accuracies of 0.83. | Varied article lengths and high computational demands for ELMo embeddings. |
| Kumar et al., 2019 | To propose and evaluate deep learning models for detecting fake news using social media and fact-checking data | Text preprocessing, tokenization, word embedding (GloVe), sentiment analysis | | Twitter and PolitiFact | Accuracy; best model achieved 88.78% | CNN + Bidirectional LSTM + Attention mechanism model outperformed others; deep learning surpassed ML models | Limited focus on source credibility; difficulty in detecting semantic drift during news propagation |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ibrishimova and Li, 2020 | To define fake news and propose a hybrid detection model using NLP and knowledge verification. | Stopword ratio, proper noun ratio, title length, ARI readability, sentiment (Google NLP API). | | Kaggle Fake News Dataset, News API, verified news outlets. | Probability scores; real news: 5–45%, fake news: 40–95%. | Hybrid model improved accuracy; combining manual and automated checks was effective. | Semantic similarity issues, new sources, and subjectivity in political content. |
| Brașoveanu & Andonie, 2020 | Improve fake news detection using semantic features and ML | Sentiment analysis, NER, relation extraction, GloVe, attention | | LIAR, PolitiFact | Accuracy (up to 53.9% on LIAR, 52.4% on PolitiFact) | Semantic features boosted model accuracy; GRU and Capsule Networks performed best | Dependency on metadata quality; domain generalization and model complexity |
| Goldani et al., 2021 | Fake news detection using capsule networks | Capsule networks, word embeddings, n-gram extraction | | ISOT, LIAR | Accuracy (99.8% ISOT, 40.9% LIAR) | Outperformed baselines; metadata improved results | Confusion between similar labels; semantic modeling needed |
| Umer et al., 2020 | Fake news stance detection using CNN-LSTM | Preprocessing, Word2Vec, PCA, Chi-square | | FNC-1 dataset | Accuracy, F1-score | PCA-based model achieved 97.8% accuracy, best among models | Data imbalance; limited to English |
| Farokhian et al., 2023 | Improve fake news detection using dual BERT (MWPBert) | Dual BERT, MaxWorth, ClaimBuster | | FakeNewsNet (PolitiFact, GossipCop) | Accuracy (85.4%), F1-score (0.64), AUC (0.863) | Dual BERT with span selection outperformed baselines | Input length limits, need for multilingual/multimodal data |
| Li et al., 2024 | Develop FactAgent for fake | GPT-3.5, expert workflow, | | Snopes, PolitiFact, | F1-score (up to 0.88), | Outperformed baselines; | Limited to titles/URLs; no full text, social, or multimodal input |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | news detection using agentic LLMs | zero-shot tools | | GossipCop | accuracy (75%–88%) | expert workflows improved results | |
| Mundra et al., 2023 | Two-phase DL-NLP framework for fake news detection | GloVe, BERT, MLP, CNN, BiLSTM, ensemble methods | | McIntyre, Kaggle, GossipCop, Politifact | F1-score (up to 95.22%) | Soft Weighted Voting Ensemble performed best | Weak on short texts; needs labeled data; no multimodal use |
| Ouassil et al., 2022 | Fake news detection using hybrid DL and word embeddings | Word2Vec, GloVe, CNN-BILSTM | | WELFake dataset | Accuracy (97.74%), F1-score, Precision, Recall | Word2Vec combo + CNN-BILSTM outperformed baselines | High dimensionality; suggested feature selection techniques |
| Agarwal et al., 2020 | Fake news detection using CNN-LSTM model | GloVe, CNN, LSTM, text preprocessing | | Kaggle Fake News Dataset | Accuracy (94.71%), Precision (97.21%) | CNN-LSTM outperformed baselines; dropout reduced overfitting | Lacked metadata use; no real-time credibility scoring |
| Hassan Ali et al., 2021 | Test fake news detectors' robustness to adversarial attacks | GloVe, CNN, RNN, MLP, Hybrid, TextAttack, LIME | | Kaggle, ISOT, LIAR datasets | Accuracy Up to **94.71%**, ASR, perturbation rate, query count | CNNs/RNNs were most robust; longer inputs helped | Accuracy-robustness trade-off; dataset bias; limited generalization |
| Faustini and Covões, 2020 | Detect fake news using language- and platform-independent features | Bag-of-Words, Word2Vec, DCDistance, custom features | | 5 datasets (Portuguese, English, Bulgarian) | Accuracy (up to 94%), F1-score | BoW best (94%); DCDistance competitive with fewer dimensions | Custom features underperformed; multilingual generalization needed |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Z. Khanam et al., 2021 | Fake news detection using ML and NLP | POS tagging, TF-IDF, tokenization | | LIAR dataset (PolitiFact) | Accuracy (>75% XGBoost, ~73% SVM & RF), Precision, Recall | XGBoost was most accurate; NLP + ML improved detection | Needed more POS-based features and further optimization |
| Prachi et al., 2022 | Fake news detection using ML, DL, and NLP techniques | Tokenization, Lemmatization, POS tagging, TF-IDF, BERT | | Kaggle fake news dataset (26,000 articles) | Accuracy: BERT 98%, LSTM 95%, Decision Tree 90%, SVM 77% | BERT outperformed all models; LSTM also highly effective | Needs topic-specific and multimodal data handling |
| Tsai, 2023 | Fake news detection using stylometric NLP and NER techniques | NER, RI/VI indexes, simulated annealing | | LIAR, Fake or Real News, FakeNews AMT | Accuracy: up to 93% (in-domain), up to 94% (cross-domain) | NER-SA outperformed ML/DL models; +50.18% gain in some domains | Overlap in entity use, no ethical/privacy handling, scalability |
| Rohit Kumar Kaliyar et al., 2020 | Fake news detection using deep CNN (FNDNet) | GloVe, CNN, ReLU, dropout | | Kaggle fake news dataset (20,800 articles) | Accuracy 98.36%, TNR 99.41%, FPR 0.59% | FNDNet outperformed ML/DL models including LSTM, CNN | Lacked multimodal/context data; future use of BERT proposed |
| Jamal Abdul Nasir et al., 2021 | Fake news detection using hybrid CNN-RNN model | GloVe, CNN, LSTM | | ISOT (45K), FA-KES (804) | Accuracy: 99% (ISOT), 60% (FA-KES) | High accuracy in-domain; weak generalization cross-domain | Overfitting; suggested dropout, multimodal, and optimization |
| Manna et al., 2020 | Profile fake news spreaders using stylometric and lexical cues | Stylometry, lexical features, emoji, clickbait terms | | PAN 2020 Twitter dataset (EN/ES, 500 users) | Accuracy: 72.5% (Spanish), 59.5% (English) | Stronger performance on Spanish tweets | Weaker stylistic cues in English tweets |

## 6. EXTRACTED STATISTICS

The Figure 1 shows that most studies in fake news detection focused on hybrid deep learning models and transformer-based approaches like BERT. Others explored semantic/stylometric methods, ensemble techniques, and language-independent systems. Overall, the trend highlights a shift toward more contextual, deep, and adaptable detection strategies

.

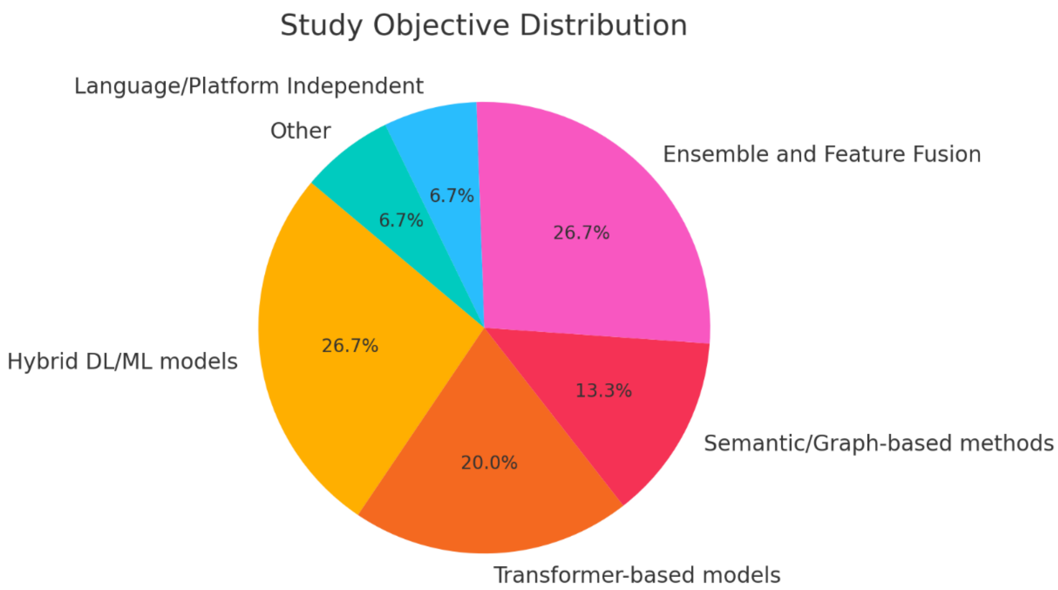


**Figure 2: Statistical representation about the Objective**

The Figure 2 shows that fake news detection studies commonly used a mix of traditional and deep learning NLP techniques. TF-IDF, GloVe, and Word2Vec were popular for text representation, while BERT, CNN, and LSTM handled contextual analysis. Other frequent methods included POS tagging, sentiment analysis, NER, and stylometry, highlighting the importance of combining linguistic and semantic features.

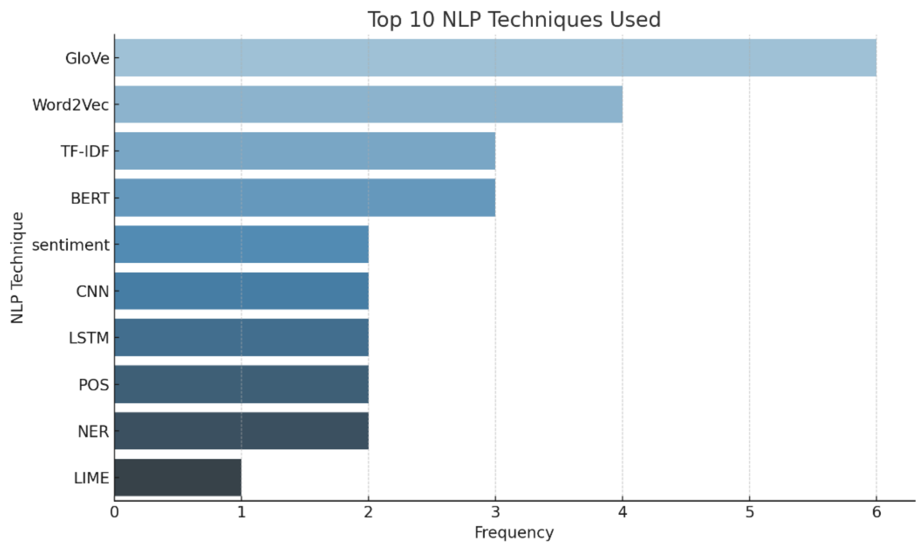


**Figure 3: Statistical representation about the Techniques Used**

The Figure 3 displaying model accuracy across fake news detection studies reveals that the majority of models achieved high accuracy, particularly within the 90–100% range. These results were largely attributed to the use of advanced deep learning architectures such as CNN-RNN hybrids, BERT, and Capsule Networks. A moderate number of models showed accuracies between 75–90%, often associated with ensemble methods or traditional NLP

techniques. Only a few studies fell below the 75% threshold, typically due to limitations in cross-domain generalization or dataset complexity. Overall, the histogram emphasizes the effectiveness of combining NLP with deep learning to enhance fake news detection performance, while also highlighting the need for improved robustness across varied contexts

.

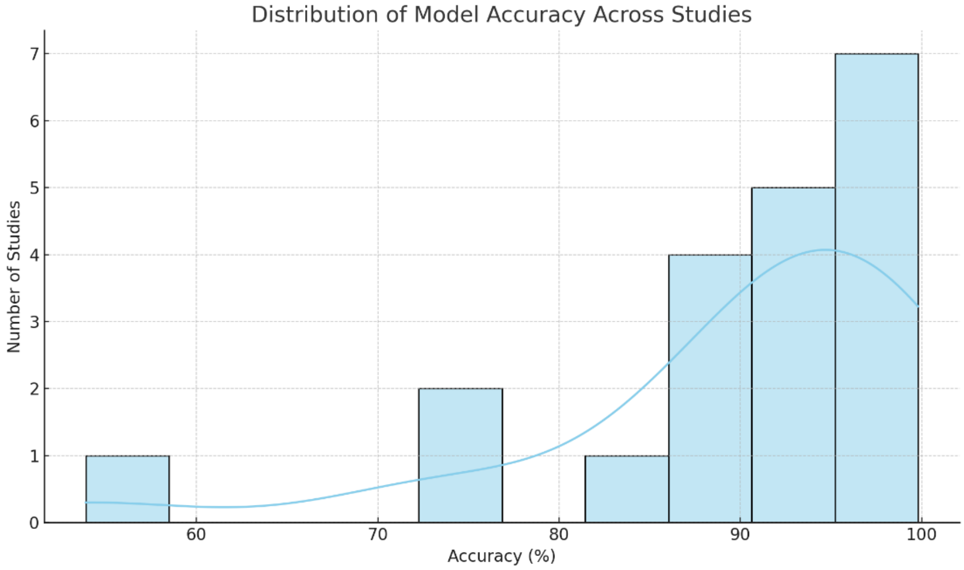


**Figure 4: Statistical representation about the Performance Metrics**

## 7. RECOMMENDATIONS

**key recommendations** for researchers and industry professionals to fully leverage NLP in combating fake news

- **Integrate Hybrid Deep Learning Architectures**
  Combine CNNs, RNNs, and attention mechanisms (e.g., CNN-BiLSTM, CNN-RNN, Transformer-based models) to enhance both semantic and sequential feature understanding.
- **Leverage Pre-trained Contextual Embeddings**
  Use state-of-the-art NLP models like **BERT**, **ELMo**, and **GPT-3.5** for better contextual analysis and to reduce reliance on manual feature engineering.
- **Utilize Stylometric and Behavioral Profiling**
  Incorporate stylometric features (e.g., punctuation, emojis, personal pronouns) to profile user intent and detect coordinated misinformation campaigns.
- **Adopt Multilingual and Cross-Domain Capabilities**
  Develop models that generalize well across **languages**, **platforms**, and **domains** using language-agnostic features and diversified datasets.
- **Incorporate Knowledge Verification Tools**
  Combine NLP with external verification mechanisms like **ClaimBuster**, **Google NLP API**, or **fact-checking APIs** to assess news credibility.
- **Apply Ensemble and Fusion Strategies**
  Use ensemble learning (e.g., Soft Voting, Bagging, Boosting) to combine the strengths of multiple classifiers and reduce overfitting.
- **Handle Imbalanced Datasets Effectively**
  Implement techniques such as **PCA**, **Chi-square selection**, and **oversampling** to improve model performance on skewed datasets.
- **Include Real-Time and Multimodal Analysis**
  Explore integrating text with **images**, **videos**, and **metadata** to better identify manipulated content across social platforms.
- **Focus on Explainability and Transparency**
  Adopt interpretable frameworks (e.g., LIME, checklists, simulated annealing) to make model decisions understandable for human evaluators.
- **Address Ethical and Privacy Concerns**
  Ensure systems are designed to avoid user profiling biases, respect privacy, and operate transparently in sensitive domains such as politics and health.

## 8. CONCLUSION

This study has highlighted the critical role of Natural Language Processing (NLP) in enhancing the accuracy, adaptability, and robustness of fake news detection systems. Through an in-depth analysis of recent research, it is evident that the integration of NLP with deep learning frameworks—particularly hybrid CNN-RNN models, BERT-based architectures, and Capsule Networks—yields superior performance in identifying fake news content. These models, often supported by advanced text representation methods like TF-IDF, GloVe, Word2Vec, and BERT embeddings, have consistently demonstrated high accuracy rates, often exceeding 90% on benchmark datasets.

Furthermore, the combination of syntactic and semantic features, such as POS tagging, named entity recognition (NER), stylometric cues, and sentiment analysis, enhances the contextual understanding of misinformation patterns. Studies also confirm the value of ensemble learning, adversarial robustness testing, and knowledge verification mechanisms in strengthening system reliability. Despite these advancements, challenges persist in achieving generalization across languages, platforms, and domains. Cross-domain transferability, ethical transparency, and real-time adaptability remain key areas for future exploration.

In conclusion, NLP continues to serve as a foundational technology in the fight against digital misinformation. Future systems should prioritize multimodal integration, multilingual support, explainable AI, and ethical safeguards to build more resilient and trustworthy fake news detection frameworks. The findings of this review reaffirm that a holistic, hybrid approach grounded in NLP can significantly contribute to curbing the spread of misinformation in the digital age.

## REFERENCES


1. Q. Su, M. Wan, X. Liu, and C. Huang, "Motivations , Methods and Metrics of Misinformation Detection : An NLP Perspective," vol. 1, pp. 1–13, 2020.
2. S. Deepak and B. Chitturi, "Deep neural approach to Fake-News identification," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2236–2243, 2020, doi: 10.1016/j.procs.2020.03.276.
3. N. R. De Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. De Medeiros, and D. M. F. Mattos, "Identifying Fake News on Social Networks Based on Natural Language Processing : Trends and Challenges," pp. 1–32, 2021.
4. M. I. Nadeem *et al.*, "SS symmetry HyproBert : A Fake News Detection Model Based on Deep Hypercontext," pp. 1–21, 2023.
5. J. Lin, G. Tremblay-taylor, G. Mou, D. You, and K. Lee, "Detecting Fake News Articles," pp. 3021–3025, 2019.
6. M. Berrondo-otermin and A. Sarasa-cabezuelo, "Application of Artificial Intelligence Techniques to Detect Fake News : A Review," 2023.
7. G. L. Ciampaglia, "Fighting fake news : a role for computational social science in the fight against digital misinformation," *J. Comput. Soc. Sci.*, 2017, doi: 10.1007/s42001-017-0005-6.
8. V. V. Hirlekar, "Natural Language Processing based Online Fake," no. Icces, pp. 748–754, 2020.
9. S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 2, pp. 1–23, 2020, doi: 10.1002/ett.3767.
10. M. Szczepański, M. Pawlicki, R. Kozik, and M. Choraś, "New explainability method for BERT - based model in fake news detection," *Sci. Rep.*, no. 0123456789, pp. 1–13, 2021, doi: 10.1038/s41598-021-03100-6.
11. A. Uppal, "Fake news detection using discourse segment structure analysis," *2020 10th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 751–756, 2020.
12. M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 6, p. 102723, 2021, doi: 10.1016/j.ipm.2021.102723.
13. R. Setiawan, V. Sagar, P. Sudhakar, and S. Mamoona, "Certain Investigation of Fake News Detection from Facebook and Twitter Using Artificial Intelligence Approach," *Wirel. Pers. Commun.*, no. 0123456789, 2021, doi: 10.1007/s11277-021-08720-9.
14. B. Probierz, P. Stefanski, and J. Kozak, "Rapid detection of fake news based on machine learning methods," *Procedia Comput. Sci.*, vol. 192, pp. 2893–2902, 2021, doi: 10.1016/j.procs.2021.09.060.
15. M. Dk *et al.*, "Exploring the Efficacy of Natural Language Processing and Supervised Learning in the Classification of Fake News Articles," 2024, doi: 10.23880/art-16000108.
16. D. A. Hasan, S. R. M. Zeebaree, M. A. M. Sadeeq, H. M. Shukur, R. R. Zebari, and A. H. Alkhayyat, "Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems - A Survey," *1st Babylon Int. Conf. Inf. Technol. Sci. 2021, BICITS 2021*, vol. 2021, no. Bicits, pp. 16–21, 2021, d oi: 10.1109/BICITS51482.2021.9509920.
17. S. Kumar, T. Roshni, and D. Himayoun, "A Comparison of Emotional Neural Network (ENN) and Artificial Neural Network (ANN) Approach for Rainfall-Runoff Modelling," *Civ. Eng. J.*, vol. 5, no. 10, pp. 2120–2130, 2019, doi: 10.28991/cej-2019-03091398.
18. M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Appl. Soft Comput.*, vol. 101, p. 106991, 2021, doi: 10.1016/j.asoc.2020.106991.
19. J. Abdul, O. Subhani, and I. Varlamis, "International Journal of Information Management Data Insights Fake news detection : A hybrid CNN-RNN based deep learning approach," vol. 1, no. December 2020, 2021, doi: 10.1016/j.jjimei.2020.100007.
20. N. R. Naredla and F. F. Adedoyin, "Detection of

hyperpartisan news articles using natural language processing technique," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, p. 100064, 2022, doi: 10.1016/j.jjimei.2022.100064.

21. T. Esther, A. K. J, P. Narayanasamy, and J. Vidya, "Attention-based C-BiLSTM for fake news detection," *Appl. Soft Comput.*, vol. 110, p. 107600, 2021, doi: 10.1016/j.asoc.2021.107600.
22. M. D. Ibrishimova and K. F. Li, *A machine learning approach to fake news detection using knowledge verification and natural language processing*, vol. 1035. Springer International Publishing, 2020. doi: 10.1007/978-3-030-29035-1_22.
23. A. M. P. Braşoveanu and R. Andonie, "Integrating Machine Learning Techniques in Semantic Fake News Detection," *Neural Process. Lett.*, vol. 53, no. 5, pp. 3055–3072, 2021, doi: 10.1007/s11063-020-10365-x.
24. M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B. W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.
25. M. Farokhian, V. Rafe, and H. Veisi, "arXiv : 2204 . 04793v2 [ cs . CL ] 17 Oct 2023 Fake News Detection using Dual BERT Deep Neural Networks," pp. 1–23, 2021.
26. X. Li and E. C. Malthouse, "Large Language Model Agent for Fake News Detection".
27. S. Mundra, J. Reddy, A. Mundra, N. Mittal, A. Vidyarthi, and D. Gupta, "An Automated Data-driven Machine Intelligence Framework for Mining Knowledge To Classify Fake News Using NLP," 2023, doi: 10.1145/3607253.
28. M. Ouassil, B. Cherradi, S. Hamida, M. Errami, and O. El, "A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model," no. January, 2022, doi: 10.14569/IJACSA.2022.0131061.
29. A. Agarwal, M. Mittal, A. Pathak, and L. Mohan, "Fake News Detection Using a Blend of Neural Networks : An Application of Deep Learning," *SN Comput. Sci.*, pp. 1–9, 2020, doi: 10.1007/s42979-020-00165-4.
30. H. Ali *et al.*, "All Your Fake Detector Are Belong to Us : Evaluating Adversarial Robustness of Fake-News Detectors Under Black-Box Settings," vol. 9, 2021, doi: 10.1109/ACCESS.2021.3085875.
31. P. Henrique, A. Faustini, and T. F. Covões, "Expert Systems with Applications Fake news detection in multiple platforms and languages," *Expert Syst. Appl.*, vol. 158, p. 113503, 2020, doi: 10.1016/j.eswa.2020.113503.
32. G. Liu, M. S. Mohamad, M. Huang, S. Pandey, S. Prabhakaran, and Z. Zhao, "Fake News Detection Using Machine Learning Approaches," 2021, doi: 10.1088/1757-899X/1099/1/012040.
33. N. N. Prachi, E. H. Rafi, E. Alam, and R. Khan, "Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms," vol. 13, no. 6, pp. 652–661, 2022, doi: 10.12720/jait.13.6.652-661.
34. C. Analysis and C. Tsai, "Stylometric Fake News Detection Based on Natural Language Processing Using Named Entity Recognition : In-Domain and," pp. 1–16, 2023.
35. E. Cuevas, "ScienceDirect FNDNet – A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, 2020, doi: 10.1016/j.cogsys.2019.12.005.
36. R. Manna, A. Pascucci, and J. Monti, "Profiling fake news spreaders through stylometry and lexical features . UniOR NLP @ PAN2020 Notebook for PAN at CLEF 2020," no. September, pp. 22–25, 2020.