

# Fake News Detection Using Machine Learning and Deep Learning Algorithms

Aman Mahendra Sahu<sup>1</sup> Samruddhi Balwant Jadhav<sup>2</sup>

<sup>1,2</sup>Student

<sup>1,2</sup>Master of Science in Information Technology

<sup>1,2</sup>Keraleeya Samajam's Model College, Thakurli, Dombivali (East), Maharashtra, India

**Abstract** — The rapid spread of misinformation and fake news across digital platforms poses significant challenges to social stability, political discourse, and cybersecurity. Traditional rule-based detection methods often fail to capture the linguistic complexity and evolving nature of fake news. This paper presents a comparative performance evaluation of multiple machine learning algorithms— including Logistic Regression, Naïve Bayes, Random Forest, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN)—for fake news detection. Using benchmark datasets, the study analyzes preprocessing techniques such as TF-IDF (Term Frequency–Inverse Document Frequency) and word embeddings, and evaluates models based on accuracy, precision, recall, F1-score, and ROC-AUC. A survey of participants was conducted to understand perceptions of fake news detection tools and the importance of algorithmic transparency. Results indicate that ensemble and deep learning approaches can provide strong predictive performance, while highlighting trade-offs between interpretability and predictive power. The findings suggest that hybrid models combining statistical and deep learning techniques may enhance detection accuracy and resilience against misinformation.

**Keywords:** Fake News Detection, Machine Learning, Deep Learning, Natural Language Processing, TF-IDF, BERT, Misinformation

## I. INTRODUCTION

The exponential growth of digital media platforms has transformed the way information is consumed and shared. While this democratization of content has enabled rapid communication, it has also facilitated the widespread dissemination of misinformation and fake news. Fake news can influence political outcomes, damage reputations, incite social unrest, and undermine trust in legitimate institutions. Detecting and mitigating fake news has therefore become a critical challenge in information technology, cybersecurity, and social sciences [1], [3], [6].

Traditional rule-based detection systems often fail to capture nuanced linguistic features and evolving strategies used by fake-news creators. As a result, machine learning (ML) algorithms have emerged as tools for automated fake-news detection. These algorithms can learn patterns from datasets, identify deceptive content, and classify news articles. However, performance varies with the algorithm, preprocessing technique, dataset, and evaluation metric [1], [3].

This study focuses on the comparative performance evaluation of Logistic Regression, Naïve Bayes, Random Forest, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN), while also discussing LSTM and BERT approaches. In addition to model characteristics, the study examines public awareness and perceptions of AI/ML-

based fake-news detection systems through a structured survey.

## II. RESEARCH GAP AND PROBLEM STATEMENT

Numerous studies have explored the application of ML and deep learning (DL) techniques for fake-news detection. Existing research has primarily focused on classification models and predictive performance using different algorithms and feature-extraction techniques. Comparatively fewer studies have examined public awareness, user trust, and perceptions regarding adoption of AI-based fake-news detection systems.

There is therefore a need to understand how users perceive the reliability, transparency, and practical usefulness of such systems in real-world environments. This study examines public perceptions through a structured survey and identifies factors that influence user confidence in AI-driven misinformation-detection solutions.

The central problem is to determine, given a piece of news content, whether it should be classified as genuine or potentially fake based on characteristics learned from previously labeled data. A machine-learning system can be represented as:

$$\rightarrow F(X) \rightarrow M(F(X)) \rightarrow Y \quad (2.1)$$

where  $X$  is the original news article or social-media post,  $F(X)$  is the extracted feature representation,  $M$  is the trained machine-learning model, and  $Y$  is the predicted class.

$$\in \{\text{Fake}, \text{Real}\} \quad (2.2)$$

More sophisticated systems can use multiple classes, such as true, false, partially true, misleading, satire, or unverifiable.

## III. OBJECTIVES AND RESEARCH QUESTIONS

The primary objectives of this study are:

- 1) Examine public awareness regarding fake news and misinformation on digital platforms.
- 2) Investigate user experiences and behaviors when encountering news content online.
- 3) Analyze perceptions toward Artificial Intelligence (AI) and Machine Learning (ML)-based fake-news detection systems.
- 4) Identify factors that influence trust in automated fakenews detection technologies.
- 5) Evaluate public opinions regarding the importance of accurate fake-news detection in society.
- 6) Study user acceptance of AI-driven tools designed to identify and reduce misinformation.
- 7) Provide recommendations for improving the effectiveness, transparency, and adoption of fake-news detection systems.

## IV. RESEARCH QUESTIONS

- 1) What is your age group?

- 2) How often do you use social media to consume news?
- 3) Have you ever encountered fake news online?
- 4) Which platform do you think contains the most fake news?
- 5) How confident are you in identifying fake news?
- 6) What do you usually do before believing in a news article?
- 7) Have you ever shared a news article that later turned out to be fake?
- 8) Which factor helps you identify fake news most effectively?
- 9) Do you think AI/ML systems can help identify fake news?
- 10) Would you recommend the use of AI-based fake-news detection tools?
- 11) What factors should an AI model consider to effectively detect fake news?

## V. PROPOSED METHODOLOGY

The present study adopts a survey-based research approach to examine public awareness of fake news and perceptions regarding AI- and ML-based fake-news detection systems. A structured questionnaire was developed and distributed through online platforms to collect responses from participants belonging to different age groups and educational backgrounds. The collected responses were organized and analyzed using descriptive statistical techniques. Frequency distributions, percentages, and graphical representations were used to identify trends in participant opinions. In addition, hypothesis testing was performed using the Chi-Square test to examine relationships between selected survey variables.

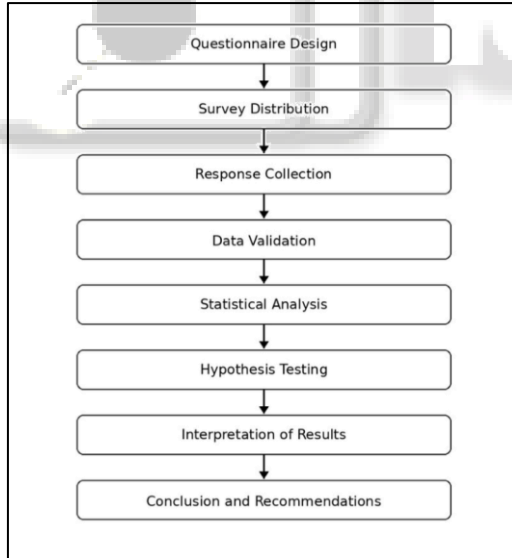


Fig. 1: Research methodology workflow.

## VI. SYSTEM ARCHITECTURE

A practical fake-news detection system can be designed as a pipeline that accepts a news article or other input, preprocesses the text, extracts numerical features, applies a trained classifier, and produces a predicted class. The architecture used as the conceptual basis of this study is shown in Fig. 2.

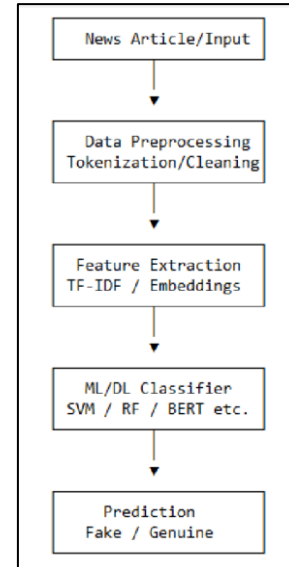


Fig. 2: Conceptual architecture for fake-news detection.

## VII. FEATURE EXTRACTION

Machine-learning algorithms require numerical representations of textual information. The following feature-extraction approaches are considered.

### A. Bag of Words

Bag-of-Words represents a document using the frequency of words appearing in the document. For example:

| Document | fake | news | election | health |
|----------|------|------|----------|--------|
| D1       | 2    | 1    | 3        | 0      |
| D2       | 0    | 2    | 0        | 4      |

Table I: Bag-Of-Words Example

Although simple, Bag-of-Words does not adequately represent word order or deeper semantic relationships.

### B. TF-IDF

Term Frequency-Inverse Document Frequency (TFIDF) assigns greater importance to words that are frequent in a particular document but relatively uncommon across the entire collection [1]. A common formulation is:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t) \quad (6.1)$$

$$\text{IDF}(t) = \log(N / \text{df}(t)) \quad (6.2)$$

Here,  $t$  denotes a term,  $d$  denotes a document,  $N$  is the number of documents, and  $\text{df}(t)$  is the number of documents containing term  $t$ . TF-IDF is widely used with traditional classifiers because it provides a relatively simple and effective representation of textual features.

### C. Word Embeddings

Word embeddings represent words as dense numerical vectors. Examples include Word2Vec, GloVe, and FastText. These representations can capture semantic relationships between words.

### D. Contextual Embeddings

Transformer models generate contextual representations in which the meaning of a word can change according to its surrounding text. BERT is a prominent example and has been investigated for fake-news classification [4].

## VIII. MACHINE LEARNING ALGORITHMS

### A. Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. The corrected probability equation is:

$$P(C | X) = [P(X | C) P(C)] / P(X) \quad (7.1)$$

It is computationally efficient and often provides a useful baseline for text classification. Its primary limitation is the assumption that features are conditionally independent given the class, which is generally not completely true for natural language.

### B. Logistic Regression

Logistic Regression estimates the probability that an input belongs to a particular class. The corrected logistic function is:

$$P(y = 1 | x) = 1 / [1 + e^{-(w^T x + b)}] \quad (7.2)$$

It is relatively simple, interpretable, and effective when combined with TF-IDF features.

### C. Support Vector Machine

Support Vector Machines attempt to find a decision boundary that separates different classes while maximizing the margin between them. SVMs can work effectively with high-dimensional sparse text representations such as TFIDF.

### D. Decision Tree

Decision Trees classify instances through a sequence of feature-based decisions. They are relatively easy to interpret but may overfit when the tree becomes excessively complex.

### E. Random Forest

Random Forest combines multiple decision trees and aggregates their predictions. Its generalization ability can be better than that of a single decision tree, particularly when the dataset contains diverse features.

### F. Passive-Aggressive Classifier

Passive-Aggressive classifiers are designed for online learning and can be useful when dealing with large streams of text data. Research comparing multiple approaches has included Passive-Aggressive classifiers alongside Naïve Bayes, Logistic Regression, Decision Trees, LSTM, and BERT.

## IX. DEEP LEARNING APPROACHES

### A. Convolutional Neural Networks

CNNs can identify local patterns in text by applying filters over sequences of words or tokens. A simplified architecture is:

Input Text → Embedding → Convolution →

Pooling → Fully Connected Layer → Classification (8.1)

CNN-based approaches have been studied in fake-news detection and can learn useful local textual patterns.

### B. Long Short-Term Memory

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network designed to handle sequential dependencies. An LSTM can process a news article as a sequence of tokens,  $x_1, x_2, x_3, \dots, x_n$ , and maintain information through hidden states. LSTMs have been used for

fake-news classification because textual order and context can contain useful signals.

### C. BERT

BERT is a transformer-based language model that produces contextual representations of text. A typical BERT-based classification architecture is:

News Article → Tokenization → BERT → Classification  
Layer → Fake/Real (8.2)

The model can be fine-tuned using a labeled fake-news dataset. Compared with traditional algorithms, BERT requires greater computational resources, but benchmark research has reported strong performance across multiple fake-news datasets [4].

## X. COMPARISON OF APPROACHES

| Approach            | Strengths                                                                                | Limitations                                          |
|---------------------|------------------------------------------------------------------------------------------|------------------------------------------------------|
| Naïve Bayes         | Fast and computationally inexpensive                                                     | Simplifying independence assumption                  |
| Logistic Regression | Simple and interpretable                                                                 | Limited representation of complex semantics          |
| SVM                 | Effective with high dimensional text                                                     | Computational cost can increase with large datasets  |
| Decision Tree       | Easy to interpret                                                                        | Can overfit                                          |
| Random Forest       | Robust ensemble method                                                                   | Less interpretable than a single tree                |
| CNN                 | Learns local textual patterns                                                            | May have limited long-range contextual understanding |
| LSTM                | Models sequential dependencies                                                           | Training can be expensive and slower                 |
| BERT                | Strong contextual language representation                                                | Requires greater computational resources             |
| Multimodal models   | Can combine text, images, and more complex data and training social context requirements |                                                      |

Table II: Comparison Of Approaches

## XI. EVALUATION METRICS

Accuracy alone is not sufficient for evaluating a fakenews detection model. The principal classification metrics are defined as follows:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (10.1)$$

$$\text{Precision} = TP / (TP + FP) \quad (10.2)$$

$$\text{Recall} = TP / (TP + FN) \quad (10.3)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (10.4)$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. For an imbalanced dataset, accuracy can be misleading; precision, recall, F1-score, and the confusion matrix should therefore also be reported.

Researchers should additionally evaluate models on datasets from different domains or time periods to determine whether the system generalizes beyond its training data.

## XII. SURVEY RESULTS AND ANALYSIS

The source document reports 31 survey records. The following charts are retained from the original survey analysis.

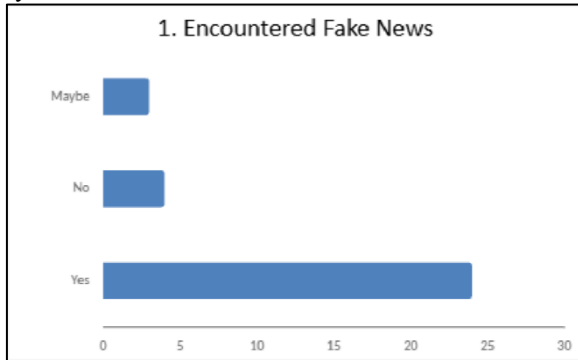


Fig. 3: Survey responses: encountered fake news.

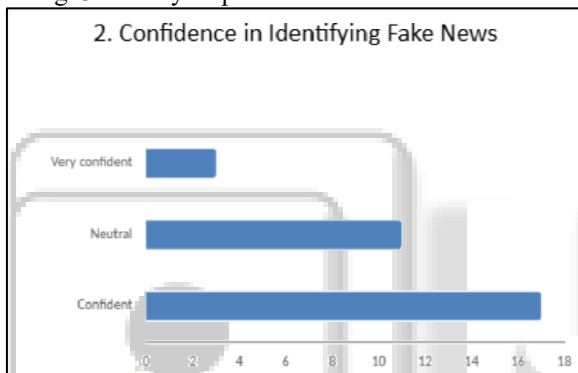


Fig. 4: Confidence in identifying fake news.

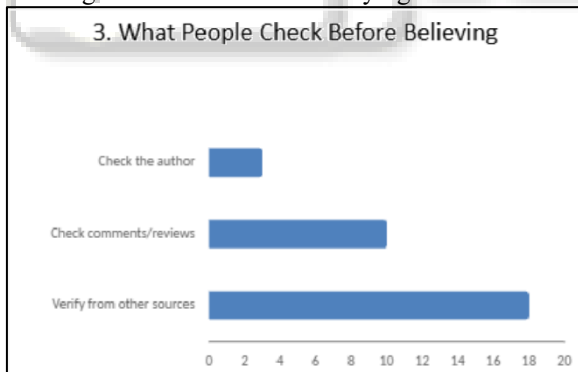


Fig. 5: Information checked before believing a news article.

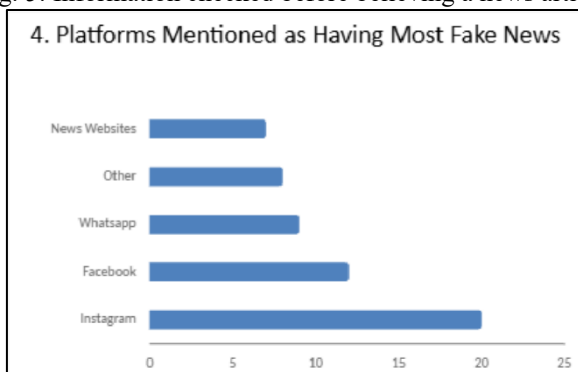


Fig. 6: Platforms mentioned as having most fake news.

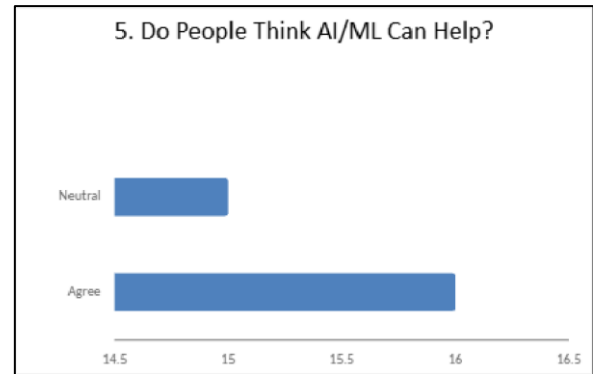


Fig. 7: Perceived ability of AI/ML to help identify fake news.

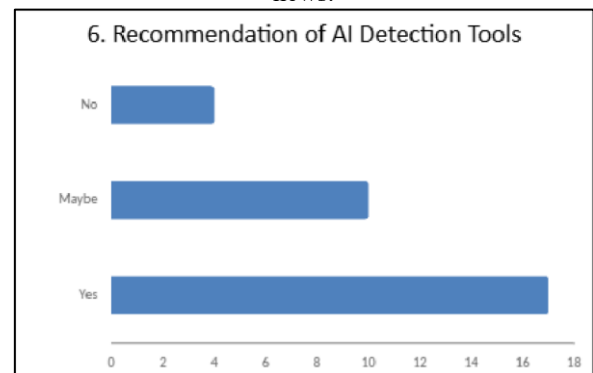


Fig. 8: Recommendation of AI-based detection tools.

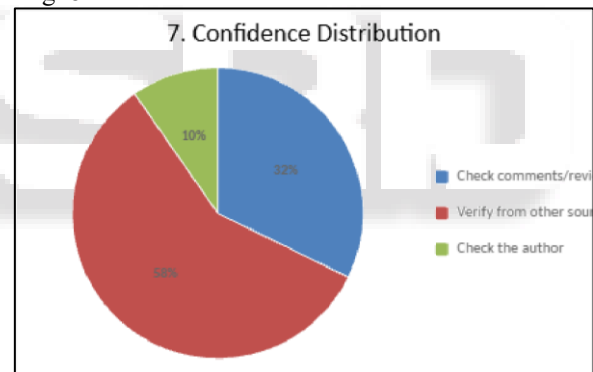


Fig. 9: Confidence distribution reported in the survey.

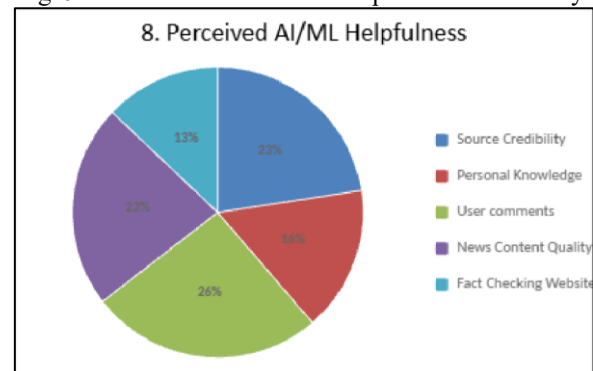


Fig. 10: Perceived AI/ML helpfulness factors

## XIII. DATA ANALYSIS AND HYPOTHESIS TESTING

A total of valid survey responses were collected from participants across different demographics. Descriptive statistical analysis was applied to evaluate perceptions of

fake-news detection systems and the comparative performance of machine-learning algorithms.

#### A. Hypothesis Statement

learning algorithms and advanced deep-learning approaches for fake-news detection.

- Alternative Hypothesis ( $H_1$ ): Advanced deep-learning approaches (e.g., CNN) provide higher detection effectiveness compared with traditional machine-learning algorithms (e.g., Logistic Regression and Naïve Bayes).

#### C. Observed Frequency Table

| Algorithm Type | Very Effective | Moderately Effective | Slightly Effective | Not Effective | Total |
|----------------|----------------|----------------------|--------------------|---------------|-------|
| Traditional ML | 6              | 10                   | 5                  | 2             | 23    |
| Deep Learning  | 12             | 8                    | 2                  | 1             | 23    |
| Total          | 18             | 18                   | 7                  | 3             | 46    |

Table III: Observed Frequency Table

#### D. Expected Frequency Table

| Algorithm Type | Very Effective | Moderately Effective | Slightly Effective | Not Effective |
|----------------|----------------|----------------------|--------------------|---------------|
| Traditional ML | 9.0            | 9.0                  | 3.5                | 1.5           |
| Deep Learning  | 9.0            | 9.0                  | 3.5                | 1.5           |

Table IV: Expected Frequency Table

#### E. Chi-Square Calculation

| Category                             | O  | E   | $(O - E)^2 / E$ |
|--------------------------------------|----|-----|-----------------|
| Traditional – Very Effective         | 6  | 9.0 | 1.0             |
| Traditional – Moderately Effective   | 10 | 9   | 0.11            |
| Traditional – Slightly Effective     | 5  | 3.5 | 0.64            |
| Traditional – Not Effective          | 2  | 1.5 | 0.17            |
| Deep Learning – Very Effective       | 12 | 9   | 1.0             |
| Deep Learning – Moderately Effective | 8  | 9.0 | 0.11            |
| Deep Learning – Slightly Effective   | 2  | 3.5 | 0.64            |
| Deep Learning – Not Effective        | 1  | 1.5 | 0.17            |

Table V: Chi-Square Calculation

Total  $\chi^2 = 3.84$ . Degrees of freedom:  $df = (2 - 1)(4 - 1) = 3$ . At  $\alpha = 0.05$ , the critical  $\chi^2$  value is 7.815.

- Null Hypothesis ( $H_0$ ): There is no significant difference in detection effectiveness between traditional machine-learning algorithms and advanced deep-learning approaches. Because the calculated  $\chi^2$  value (3.84) is below the critical value (7.815), the null hypothesis is not rejected at the 0.05 significance level. The observed distribution contains differences between the two groups, but the reported sample does not provide sufficient evidence of a statistically significant association at this threshold.

The descriptive distribution shows more “Very Effective” responses for deep learning and more “Slightly Effective” and “Not Effective” responses for traditional ML. These descriptive patterns should not be interpreted as proof of superiority because the chi-square test did not reach statistical significance at  $\alpha = 0.05$ .

### XIV. CHALLENGES AND LIMITATIONS

#### A. Dataset Bias

Many datasets are concentrated around particular topics, languages, countries, or political contexts. Consequently, a model may learn characteristics of a particular dataset rather than general principles of false information. Dataset

#### B. Selection of Variables

For hypothesis testing, the following categorical variables were selected:

- Variable 1 (Independent): Type of algorithm used (Traditional ML vs. Deep Learning).
- Variable 2 (Dependent): Detection effectiveness (Very Effective / Moderately Effective / Slightly Effective / Not Effective).

availability and concentration in English and political news remain limitations [3].

#### B. Linguistic Complexity

Fake news can deliberately imitate legitimate journalism. A fabricated article may therefore have grammatically correct language and a professional writing style.

#### C. Sarcasm and Satire

Satirical articles can resemble false information even though their purpose is entertainment or commentary. A purely text-based classifier may have difficulty distinguishing satire from intentional deception.

#### D. Changing Topics

A model trained on old news may not perform well on new events because vocabulary, entities, narratives, and writing styles change over time.

#### E. Multimodal Information

Modern misinformation frequently combines text with photographs, videos, graphics, and manipulated media. Text-only models therefore cannot capture every relevant signal.

#### F. Explainability

A prediction such as “Fake: 94%” does not necessarily explain why the system reached that conclusion. For high-impact applications, users may need interpretable evidence rather than only a classification label.

#### G. False Positives and False Negatives

A detection system can incorrectly classify legitimate information as fake or fail to detect fabricated information. Automated predictions should therefore be treated as signals for further verification rather than unquestionable judgments.

### XV. FUTURE SCOPE

#### A. Multimodal Detection

Future systems can combine textual content, images, videos, user interactions, network structure, and metadata to provide a more comprehensive representation of online information.

#### B. Graph Neural Networks

News articles, users, publishers, and social interactions can be represented as graphs. Graph neural networks can then learn relationships between these entities.

#### C. Retrieval-Augmented Fact Verification

Instead of relying exclusively on patterns learned from training data, future systems can retrieve information from authoritative sources and compare claims against external evidence.

#### D. Explainable AI

A useful system should potentially identify the evidence or linguistic/contextual characteristics that contributed to its prediction.

#### E. Cross-Domain Generalization

Researchers should test whether models trained on political news can work on health, finance, science, or other domains.

#### F. Multilingual Detection

*Much existing research is concentrated on English. Developing reliable models for Indian and other languages is an important research direction. Robustness Against AI-Generated Content*

Generative AI can produce increasingly convincing fabricated text. Future fake-news research will need to consider content generated or substantially modified by large language models, while avoiding the assumption that AI-generated content is inherently false.

### XVI. CONCLUSION

Fake-news detection is an important application of machine learning and natural language processing. Traditional algorithms such as Naïve Bayes, Logistic Regression, SVM, Decision Trees, and Random Forests provide useful and computationally efficient baselines. Deep-learning approaches such as CNNs and LSTMs can learn more complex patterns, while transformer-based approaches such as BERT provide contextual representations [4].

However, high performance on a benchmark dataset does not necessarily mean that a model can reliably determine

truth in the real world. Dataset bias, domain shifts, evolving narratives, multilingual content, satire, multimodal misinformation, and the need for explainable evidence remain significant challenges. Research has therefore increasingly moved toward combining textual information with social context, external knowledge, graph structures, and multimodal data [3], [9].

Overall, machine learning can serve as an effective component of a broader information-verification system. A robust future system can combine automated detection with evidence retrieval, source analysis, human review, and transparent explanations rather than treating a model’s classification as definitive proof.

### 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 Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- [2] H. Ahmed, I. Traore, and S. Saad, “Detecting opinion spams and fake news using text classification,” *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.
- [3] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories, detection methods, and opportunities,” *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, 2020.
- [4] R. K. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,” *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 13475–13488, 2021.
- [5] W. Y. Wang, “‘Liar, liar pants on fire’: A new benchmark dataset for fake news detection,” in *Proc. 55th Annual Meeting of the Association for Computational Linguistics*, pp. 422–426, 2017.
- [6] X. Zhang and A. A. Ghorbani, “An overview of online fake news: Characterization, detection, and discussion,” *Information Processing & Management*, vol. 57, no. 2, p. 102025, 2020.
- [7] N. J. Conroy, V. L. Rubin, and Y. Chen, “Automatic deception detection: Methods for finding fake news,” in *Proc. Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.
- [8] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in *Proc. 2017 ACM Conf. Information and Knowledge Management*, pp. 797–806, 2017.
- [9] T. Zhou, J. Wu, and R. Zafarani, “Fake news detection: A survey of graph-based approaches,” *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 5, pp. 1–37, 2021.
- [10] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.