

## ANALYSIS OF SCIENTIFIC APPROACHES TO ENSURE HIGH ACCURACY IN FAKE NEWS DETECTION

1 Uzoqov Lochinbek Mamurjon o'g'li

2 Qurbonov Behzod Bahodir o'g'li

1Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Doctoral Student, 2PDP University, Senior Lecturer  
lochinbekuzoqov97@gmail.com

### Abstract

This article aims to assess the relevance and importance of scientific research and research aimed at improving fake news detection algorithms based on machine learning, deep learning, transformer models and social network analysis. The article assesses the strengths and weaknesses of each research paper and provides recommendations for effective approaches. In addition, by combining scientific approaches, the possibilities of achieving high accuracy and reliable results in identifying fake news are considered.

**Keywords:** Fake news, machine learning, deep learning, transformer models, social network analysis, fake news detection algorithms, hybrid models, explainable AI.

### Introduction

In recent years, fake news has become one of the serious problems on a global scale. They can lead to the spread of incorrect information in the society and disrupt social, political and economic stability. The rapid spread of fake news through the Internet and social media causes people to believe fake news. Therefore, there is a growing need to develop effective algorithms and methods for detecting fake news. By studying the effectiveness of different techniques in detecting fake news, there are opportunities to further develop this field and develop more sophisticated models in the future. This paper uses approaches based on machine learning, deep learning, and social network analysis to identify various advances in fake news and constraints are evaluated.

### 2 MATERIALS AND RESEARCH METHODS

#### "Deep Learning-Based Fake News Detection".

Different machine learning models, in particular deep learning methods, can be used to identify fake news. Working with many datasets, the effectiveness of each model is evaluated. Models such as BERT, RNN and CNN are compared and their role in detecting fake news is studied.

---

Advantages and disadvantages of deep learning methods for detecting fake messages:

Advantages:

- **Effective use of deep learning models:** The research shows the effectiveness of deep learning models such as BERT in detecting fake news. BERT deeply learns complex contextual features through bidirectional text analysis, and it shows high accuracy, especially in identifying fake news.
- **Work with a wide range of datasets:** This research has been studied with several different datasets, which allows the generation of models. It makes it possible to test the stability of machine learning models through different datasets [1].

Disadvantages:

- **Scarcity of resources:** Training deep learning models requires a large amount of data and resources. Training a BERT model takes a lot of time, which makes it difficult to use in practice.
- **Complexity of real-time operation:** Due to the complexity of deep learning models, they can be difficult to use in real-time news analysis. This limits the ability to immediately detect fake news.

### **"Hybrid Deep Learning for Fake News Detection".**

In this research, an approach combining deep learning hybrid models, including CNN and RNN models, was used, and a model aimed at increasing the effectiveness of fake news classification was developed.

Advantages:

- **Hybrid approach:** In this study, a hybrid approach combining CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) models was used. While CNN models are powerful in analyzing visual data and text, RNN models excel in learning time sequences. Together, this increases accuracy and sensitivity in detecting fake news.
- **Deep learning of features:** Hybrid models deeply analyze various features of the text, which gives more accurate results in identifying fake news. By analyzing text and other content together, different messages can be effectively identified.

Disadvantages:

- **Complexity of combination:** Combinations between CNN and RNN models are complex and require a lot of time and resources to train and implement. This combined approach can make real-time performance difficult.

**Difficulty interpreting models:** Interpreting the results of hybrid models, especially deep learning models, can be difficult. It is difficult to explain why a model determines whether it is fake or real, which increases the need for Explainable AI approaches.

---

**"Machine Learning for Detection of Fake News" (MIT, 2020).**

Algorithms developed on the basis of specific stylometric properties and textual analysis are proposed in this analysis. The main focus is on determining the fakeness of the message through the connections between words and the stylistic features of the text.

**Advantages:**

- Simplicity and effectiveness of machine learning models: This thesis examines how effective machine learning models such as Logistic Regression, Naïve Bayes, and Random Forest are in detecting fake news. These models are simple and require few resources, but achieve relatively good results[2].
- Fast and low-resource approach: The unique advantage of machine learning models is that they are fast and do not require a lot of information and resources. The models presented in this dissertation work effectively with small and medium-sized datasets.

**Disadvantages:**

- Limited capacity for deep learning of complex texts: Traditional machine learning methods are limited in learning deep contextual meanings, and their accuracy decreases when working with complex texts.
- Difficult to analyze contextual links: Machine learning models have little ability to accurately distinguish the contextual links of the text, which leads to limitations in identifying fake news that requires deep analysis.

**"Fake News Detection Tools and Methods: A Review (arXiv preprint)".**

In this method, the most modern methods used to detect fake news, as well as real-time tools, are analyzed. The study examines the tools available for working with different datasets and algorithms used to detect fake news. The article also provides detailed information on automatic detection methods and techniques based on content and social context.

**Advantages:**

- Extensive analysis: This article compares the most modern tools and methods used to detect fake news. This analysis shows the performance of different algorithms, approaches and different datasets.
- Show a wide range of technical tools: The article analyzes the use of real-time tools, in particular, machine learning models. Different methods are compared to each other to determine which approaches are more effective [2].

**Disadvantages**

- Difficulty of practical application: Since the article contains more theoretical analysis, the practical application of some technologies can be very complex and resource-intensive. Also, it can be difficult to test each of the technical tools in a practical application.

- Require flexibility: Many of the tools and methods presented in this study are not always flexible to different types of datasets, meaning that some methods only work well with datasets from a specific domain or language.

The above four methods and two articles have made progress in detecting fake news by applying various effective technologies. Although Deep Learning and Hybrid Models provide high accuracy and sensitivity, they require a lot of resources and time, which limits their practical application. On the other hand, Traditional Machine Learning Models are characterized by their simplicity and speed, but their effectiveness in analyzing complex texts is lower. Although real-time analysis of social media news, hybrid approaches, and benchmarking techniques have greatly improved results, these approaches have limitations such as resource scarcity and lack of interpretability.

### **3 STRATEGIES TO IMPROVE THE EFFECTIVENESS OF DETECTING FAKE NEWS**

#### **1. Better feature extraction with advanced NLP models**

- Application of transformer models: models such as BERT and RoBERTa are achieving high results in identifying fake news by understanding the relationship between words in a dual way. These models clearly extract the contextual meaning of complex data.
- Multimodal analysis: Fake news often includes text, images, and videos. Using models such as CLIP, the results can be made more accurate by combining text and image analysis [3].

#### **2. Verification using fact-checking tools and knowledge graphs**

- Integration with Knowledge Graphs: Knowledge graphs can be used to identify fake news by comparing the facts in an article with trusted sources. For example, events and facts in the article are checked against sources such as Wikipedia.
- Real-time fact-checking APIs: By integrating tools such as Google Fact Check Explorer to continuously check news, the veracity of news is verified.

#### **3. Combining machine learning and social network analysis**

- Source Credibility Analysis: By combining content analysis with social media source credibility, news can be more accurately identified. Fake news often spreads through unreliable sources.
- Analysis of user behavior: fake news can be detected at an early stage by monitoring users' reactions to news in social networks (like, share).

#### **4. Hybrid models (combining different approaches)**

- Ensemble learning: Efficiency can be increased by combining different models. For example, by extracting the BERT features, the classification process can be optimized using the Logistic Regression model.

---

Which method is more effective?

The most effective method for detecting fake news depends on the situation and the dataset. In general, transformer models (BERT and RoBERTa) show high accuracy, especially when working with large amounts of data. Hybrid models combine deep learning methods and are effective in performing high-level complex analyses. Traditional machine learning models are more resource-efficient and faster when working with small datasets, but they don't perform well when deeper analysis is required[3].

Different approaches are often combined to achieve superior results, for example, efficiency is increased by combining transformer models and social network diffusion analysis.

Term Frequency – TF. It measures how often words are used in a text, which helps detect fake news, especially with emotional or sensational words.

$$TF(w) = \frac{f(w)}{N} \quad (1)$$

where  $f(w)$  is the total number of occurrences of the word  $w$  in the text, and  $N$  is the total number of words in the text.

Stylometric Analysis

Stylometric analysis analyzes sentence structures, punctuation, and emotional words.

Statistical tests (such as the chi-square test) can be used to measure stylistic differences[4]:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (2)$$

Here  $O$  and  $E$  are the expected and observed frequencies.

N-gram Models

Learns specific features of a text based on a sequence of words or characters. Fake messages often contain emotional or sensational phrases[5]. For example, the word sequence probability for bigrams is calculated as follows

$$p(\omega_n|\omega_{n-1}) = \frac{\text{Count}(\omega_{n-1},\omega_n)}{\text{Count}(\omega_{n-1})} \quad (3)$$

Here,  $\omega_n$  –the current word,  $\omega_{n-1}$  – the previous word.

Sentiment Analysis

Focuses on identifying the emotion in the text. Fake messages can often have a negative or threatening tone.

Overall emotional assessment of the text:

$$\text{Sentiment Score} = \frac{1}{n} \sum_{i=1}^n S(\omega_i) \quad (4)$$

Here  $S(\omega_i)$  is the sentiment value (positive, negative or neutral) of the word  $\omega_i$ , and  $n$  is the number of words in the text.

---

### Transformer Models (BERT, RoBERTa)

BERT and RoBERTa models use the self-attention mechanism and allow to understand the context of words in the text in two directions.

To classify a message as fake or real, the BERT result can be evaluated through the classification layer[6]:

$$\text{output} = \text{softmax}(W \cdot h_{[\text{CLS}]} + b) \quad (5)$$

$h_{[\text{CLS}]}$  is a representation of the token, and  $W$  and  $b$  are parameters that can be studied.

### 4 Conclusion

In the scientific works and dissertations mentioned above, the approaches to identifying fake news using different technologies and algorithms are analyzed. Among its achievements are the high accuracy of machine learning and deep learning models and the ability to work with new technologies, while its shortcomings are related to the lack of resources and difficulties in obtaining interpretable results. Hybrid models may be one of the most effective technologies for detecting fake news in the future, but their practical application requires resources and technical skills.

### References

1. Ting Su (2022). Automatic Fake News Detection on Twitter. University of Glasgow, PhD dissertation.
2. Nanyang Technological University (2022). Deep Learning-Based Fake News Detection.
3. Auburn University (2022). Fake News Detection: Engagement and Propagation Path, PhD dissertation.
4. Massachusetts Institute of Technology (MIT) (2020). Machine Learning for Detection of Fake News.
5. Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research, Detection Methods, and Opportunities. arXiv preprint arXiv:1812.00315.
6. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. doi:10.1145/3137597.3137600.