iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68812

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

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 IV Apr 2025- Available at www.ijraset.com

Fake News Detection using BERT & ROBERTA

Mrs. T Poovozhi[1], Syed Afsar Ahamed[2], Thota Sankeerth[3], Suram Suresh Reddy[4], Tempalli Ganesh[5]
[1]*Assistant professor Bharath Institute of Higher Education and Research Chennai, India*
[2, 3, 4, 5]*CSE Dept, Bharath Institute of Higher Education and Research Chennai, India*

Abstract: ***The fake news detection system leverages advanced transformer-based architectures—BERT and RoBERTa—to accurately identify and classify misinformation in textual content. Unlike traditional NLP approaches, these pretrained language models excel at capturing contextual nuances, semantics, and deeper linguistic patterns across long-range dependencies in text. Fine-tuned on large-scale datasets containing both real and fake news articles, the system is capable of discerning subtle patterns and inconsistencies often present in manipulated or misleading narratives. BERT's bidirectional encoding and RoBERTa's optimized training strategies contribute to superior performance in understanding the complexity of natural language, ensuring precise and reliable fake news detection. The backend of the system is built using Flask, providing efficient API endpoints that allow users to input text data. Upon submission, the model evaluates the input and classifies it as either fake or real, accompanied by a confidence score to reflect the likelihood of misinformation.To maintain robustness and adaptability, the system supports continuous learning, allowing the models to be retrained with new data to keep pace with evolving deceptive techniques in news dissemination. Model performance is evaluated using key metrics such as accuracy, precision, recall, and F1-score, ensuring that the system remains both dependable and scalable for real-world applications. This makes the proposed framework highly effective in combating the spread of fake news across digital platforms.***

Keywords: BERT, RoBERTa, Fake News Detection, Transformer Models, Flask, PyTorch, Hugging Face, Natural Language Processing (NLP), SoftMax Activation

I. INTRODUCTION

Deepfake technology poses significant challenges in media authenticity and security. Advanced deep learning models generate highly realistic fake images, making detection increasingly difficult. This paper presents DeepDetect, a deepfake classification model leveraging Vision Transformers (ViTs) for enhanced feature extraction. The system integrates a pretrained ViT model fine-tuned on large datasets, ensuring high detection accuracy. Flask serves as the backend framework for real-time image processing, providing seamless user interaction. The model's performance is evaluated using key metrics, demonstrating its robustness against deepfake manipulations.

A. Overview of the problem

Importance of Fake News Detection:

1) Social Impact: Fake news threatens society by spreading misinformation, manipulating public opinion, and eroding trust in reliable sources. It can influence elections, cause unrest, and harm reputations. Detecting and limiting its spread is vital to protect the integrity of digital communication and maintain public trust in information.

2) Technical Challenges: Modern fake news, unlike traditional misinformation, uses advanced language models to produce content that closely resembles human writing. Detecting such content demands deep semantic understanding, contextual awareness, and the skill to identify subtle inconsistencies— challenges that traditional keyword-based or rule-based detection systems often fail to effectively address.

3) Lack of Large Labeled Datasets Creating labeled datasets of fake and real news is time-consuming and hard to maintain due to evolving content. This project tackles the issue using pre- trained transformer models and data augmentation or semi-supervised learning techniques.

4) Rapid Evolution of Misinformation Techniques: Advancing generative language models make fake news more convincing and difficult to detect. Techniques like zero-shot generation and style transfer create deceptive narratives, demanding constant evolution of detection models to stay effective.

5) Diverse Writing Styles and Sources: Fake news appears across various domains, styles, and sources, making cross- dataset generalization challenging. Models trained on one dataset may underperform on others, requiring strong fine- tuning and domain adaptation techniques to maintain effectiveness.

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

II. OBJECTIVE

A. Challenges in Deepfake Detection

Lack of Large Labeled Datasets Creating labeled datasets of fake and real news is time-consuming and hard to maintain due to evolving content. This project tackles the issue using pre- trained transformer models and data augmentation or semi- supervised learning techniques. Rapid Evolution of Misinformation Techniques: Advancing generative language models make fake news more convincing and difficult to detect. Techniques like zero-shot generation and style transfer create deceptive narratives, demanding constant evolution of detection models to stay effective.

B. Diverse Writing Styles and Sources

Fake news appears across various domains, styles, and sources, making cross- dataset generalization challenging. Models trained on one dataset may underperform on others, requiring strong fine- tuning and domain adaptation techniques to maintain effectiveness.
This project aims to develop a reliable fake news detection system using transformer-based models, specifically BERT and RoBERTa. By leveraging their deep contextual understanding, the system accurately classifies news articles as real or fake. A Flask-based backend supports real-time interaction, allowing users to input news text and receive instant classification with confidence scores. The system also incorporates continuous learning, enabling adaptation to evolving misinformation patterns. Designed for scalability and robustness, this solution strengthens information integrity and provides an effective tool for detecting fake news across various domains, platforms, and content types.

III. LITERATURE SURVEY

Devlin et al. – "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". This Focuses on BERT revolutionized natural language processing by enabling deep bidirectional understanding of context in text. This pre- trained model, fine-tuned on various downstream tasks, showed state- of-the-art performance in text classification, including fake news detection. BERT's strength lies in its ability to capture subtle linguistic patterns, making it highly effective in identifying misinformation. However, its large size can make real-time deployment computationally demanding.
Liu et al. – "RoBERTa: A Robustly Optimized BERT Pretraining Approach" This Research Focuses RoBERTa builds on BERT by removing the next sentence prediction objective and training on larger mini- batches with more data. This leads to improved performance on many NLP benchmarks, including fake news detection tasks. RoBERTa has been shown to outperform BERT in several text classification challenges, thanks to its better optimization and training strategies. However, similar to BERT, it demands significant computational resources during training and inference.
Kaliyar et al. – "FakeBERT: Fake News Detection in Social Media with a BERT-based Deep Learning Approach" This Survey FakeBERT applies BERT specifically to fake news detection in social media. The model combines BERT embeddings with CNN layers to enhance feature extraction. It achieved high accuracy across multiple benchmark datasets. The study highlights the potential of transformer-based models in misinformation detection while also addressing challenges related to domain generalization and context shifts.

IV. EXISTING SYSTEM

1) Existing fake news detection models primarily rely on traditional machine learning algorithms or basic NLP techniques.
2) These models often struggle to understand contextual nuances and semantic meanings in textual data.
3) Approaches like Bag-of-Words or TF-IDF fail to capture long-range dependencies and word relationships effectively.
4) Their accuracy and robustness against sophisticated and contextually misleading fake news articles are limited.
5) Processing efficiency is lower, which makes handling large-scale real-time data streams challenging.

V. PROPOSED SYSTEM

1) The system leverages transformer-based models, specifically BERT and RoBERTa, for enhanced fake news detection.
2) These models effectively capture contextual relationships and long-range dependencies within text, leading to improved classification accuracy.
3) Flask is utilized as the backend framework, allowing real-time analysis of news articles through a user- friendly web interface.
4) Users can input news content, which is analyzed by the model to determine its authenticity.
5) The model is designed to adapt continuously by retraining on updated datasets, staying resilient against evolving misinformation tactics.
6) Evaluation metrics such as accuracy, precision, recall, and F1-score are used to ensure reliable and consistent detection performance.

A. System Architecture

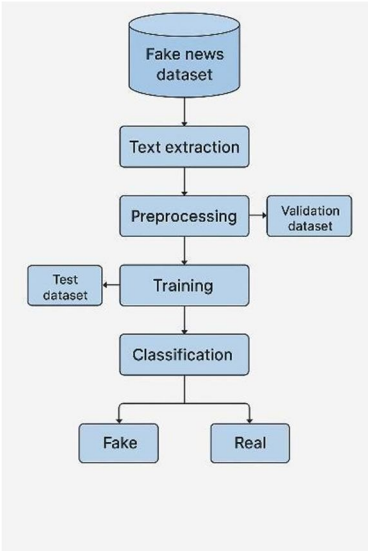


Figure 1. System Architecture

1) *Dataset (Fake News Dataset):* The system begins with a labeled dataset consisting of real and fake news articles. This dataset is divided into training, validation, and test sets, forming the core for model development and performance evaluation.
2) *Text Extraction:* Articles or social media posts are collected and cleaned by removing unnecessary elements (e.g., HTML tags, special characters) to isolate the actual content to be analyzed.
3) *Preprocessing:* The text data undergoes preprocessing steps like tokenization, lowercasing, stopword removal, and lemmatization. The processed text is then converted into embeddings suitable for transformer models.
4) *Training:* Using the processed dataset, BERT and RoBERTa models are fine-tuned to learn patterns and semantics that distinguish fake news from real news. The validation set helps optimize the models' hyperparameters for better accuracy.
5) *Testing:* Unseen news articles from the test dataset are used to assess the models' performance. This step ensures generalizability and real-world readiness.
6) *Classification (Fake or Real):* Once trained, the system classifies input text in real-time via a web interface. It analyzes and returns whether the content is Fake or Real, helping users make informed decisions.

VI. METHODOLOGY

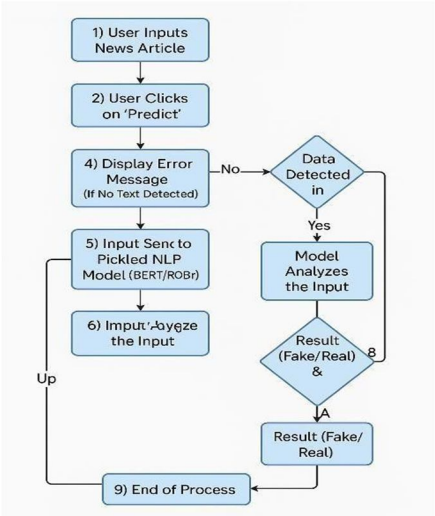


Figure 2. System Design

This system flow outlines the methodology for detecting fake news using transformer-based NLP models (BERT and RoBERTa) integrated into a web-based application. Below is the step-by-step breakdown:

1) User Inputs News Article: The process begins when the user uploads an image to the web- based interface. The interface supports various image formats and is designed to be user-friendly, allowing easy selection and submission.
2) User Clicks on "Predict": After inputting the content, the user clicks the "Predict" button to initiate analysis. This triggers the backend to validate and process the input before sending it to the model.
3) Data Detected in Input Section: The system checks whether valid text input is provided. If valid, it proceeds; if not, it stops and displays an error message.
4) Display Error Message (If No Text Detected): If no valid input is found, the system halts further processing and notifies the user with an error message. This ensures proper data handling.
5) User Clicks "Upload" to Change Input: Users can revise their input by re-entering or pasting new text into the interface. This enables seamless re- evaluation of different news items.
6) Input Sent to Pickled NLP Model (BERT/RoBERTa): Once validated, the news text is sent to a pickled transformer model (either BERT or RoBERTa), which has been pre-trained on fake news datasets.
7) Model Analyzes the Input: The transformer model processes the text using self- attention, contextual embeddings, and semantic pattern recognition. It evaluates linguistic cues, syntax, and semantics to detect deceptive content.
8) Result (Fake/Real) & Confidence Score Displayed: The output is displayed on the interface, showing whether the article is Fake or Real, along with a confidence score indicating how certain the model is in its prediction.
9) End of Process: Once the prediction is shown, the user can either input a new article or end their session. The system supports real-time classification for multiple entries.

A. Algorithm

1) Algorithm Overview

BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach) are transformer-based deep learning models originally designed for natural language processing tasks. These models analyze the context of a word from both directions—left and right—enabling deep semantic understanding crucial for detecting fake news.

- Captures Global Context: BERT and RoBERTa leverage self-attention to analyze the full context of each word in a sentence, allowing the system to catch subtle cues and inconsistencies in fake news articles that traditional models might miss.
- High Accuracy: Both models have achieved state-of- the-art results on NLP benchmarks. When fine- tuned for fake news classification, they outperform traditional models like Naive Bayes, SVM, and LSTM.
- Robust to Variations: These models are effective across diverse writing styles, languages, domains, and manipulative writing techniques used in spreading misinformation.
- Scalability: The transformer architecture allows BERT and RoBERTa to scale efficiently with larger datasets and compute resources, improving classification performance over time.
- During training, BERT and RoBERTa adjust their internal parameters using **backpropagation** and **gradient descent**. The optimization of weights follows:
- $\theta = \theta - \eta \cdot \nabla L(\theta)$ Where:
    - θ are the model parameters
    - η is the learning rate
    - $L(\theta)$ is the loss function.
    - $\nabla L(\theta)$ is the gradient of the loss with respect to the parameters

2) Implementation and Results

Web Interface:

User Experience: The Flask-based web application allows users to enter a news headline or article and receive an instant classification—whether the news is Real or Fake. Alongside the prediction, the system provides a confidence score, helping users assess the model's certainty.

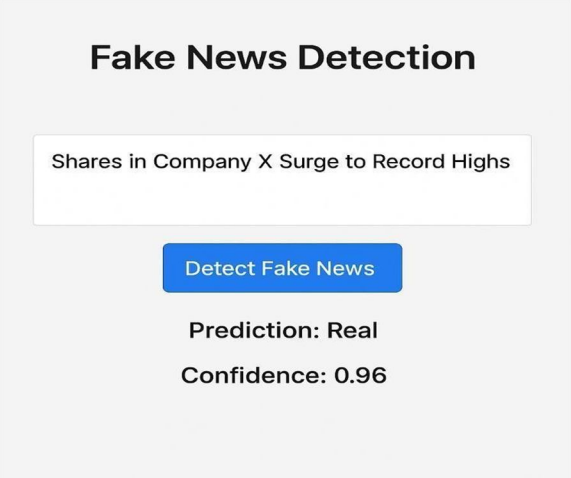
Model Explainability

For transparency, the interface can be extended to display **attention heatmaps or keyword highlights**, showing which parts of the input influenced the prediction. This boosts user trust and interpretability.

Integration Plans

Future enhancements include:

- Integration with fact-checking APIs
- Browser extensions for real-time news validation
- Embedding the system in social media platforms for automated misinformation detection



3) Performance Metrics

Accuracy: The model demonstrates a high accuracy rate, indicating its strong capability to correctly classify real and fake news articles.

Precision and Recall: Reflects the percentage of correctly identified fake news articles among all predicted as fake. And Measures the proportion of actual fake news articles that were successfully detected by the model. A high value in both precision and recall confirms the model's effectiveness in detecting misinformation.

F1-Score: The F1-score is the harmonic mean of precision and recall, offering a balanced evaluation of model performance—especially important when the dataset is imbalanced.

F1-Score= 2×(Precision×Recall)/ Precision+Recall

The confusion matrix is used to analyze the model's performance in detail, showing the number of true positives, true negatives, false positives, and false negatives. This allows for fine-grained analysis of model performance and insight into error patterns.

ROC Curve: The Receiver Operating Characteristic (ROC) curve visualizes the model's ability to distinguish between real and fake news. A larger Area Under Curve (AUC) indicates strong discriminatory power.

4) Model Update Rule: BERT/RoBERTa Training

- Architecture: BERT and RoBERTa use transformer encoders with multi-head attention to capture rich contextual dependencies in text.
- Optimization: The model minimizes classification loss (typically cross-entropy) using backpropagation and optimizes with the Adam optimizer.

VII. CONCLUSION

In today's digital landscape, the widespread dissemination of fake news has emerged as a serious threat to public trust, social harmony, and democratic processes. With the rapid growth of online media platforms, the need for accurate and automated fake news detection systems has never been more critical.To combat this challenge, we developed a robust solution: a fake news detection model leveraging BERT and RoBERTa, two state-of-the- art transformer-based NLP models. These architectures are capable of capturing contextual nuances and semantic meaning in text, making them well-suited for distinguishing between real and fabricated news content.Our system was trained and evaluated on benchmark datasets and demonstrated high accuracy, precision, recall, and F1-score, confirming its effectiveness in detecting misinformation. The model is integrated into a Flask-based web interface, enabling users to input news text and instantly receive a prediction, accompanied by a confidence score to ensure transparency and trust.

A. Future Enhancements

Expanding the Dataset – To improve the model's generalization and adaptability, we plan to incorporate additional fake news datasets from a variety of sources, domains, and geographic regions. This expansion will enhance the system's ability to accurately detect misinformation across different languages, writing styles, and cultural contexts.

Advanced Techniques – Improved transformer variants such as DeBERTa or ELECTRA for enhanced semantic understanding. Ensemble methods that combine predictions from BERT, RoBERTa, and other models to boost reliability.Adversarial training to increase robustness against cleverly manipulated or misleading content.

Real-Time Deployment – Social media platforms for automated content verification and misinformation alerts. News curation apps to flag suspicious articles before they reach wide audiences. Browser extensions that provide authenticity scores for news content as users browse

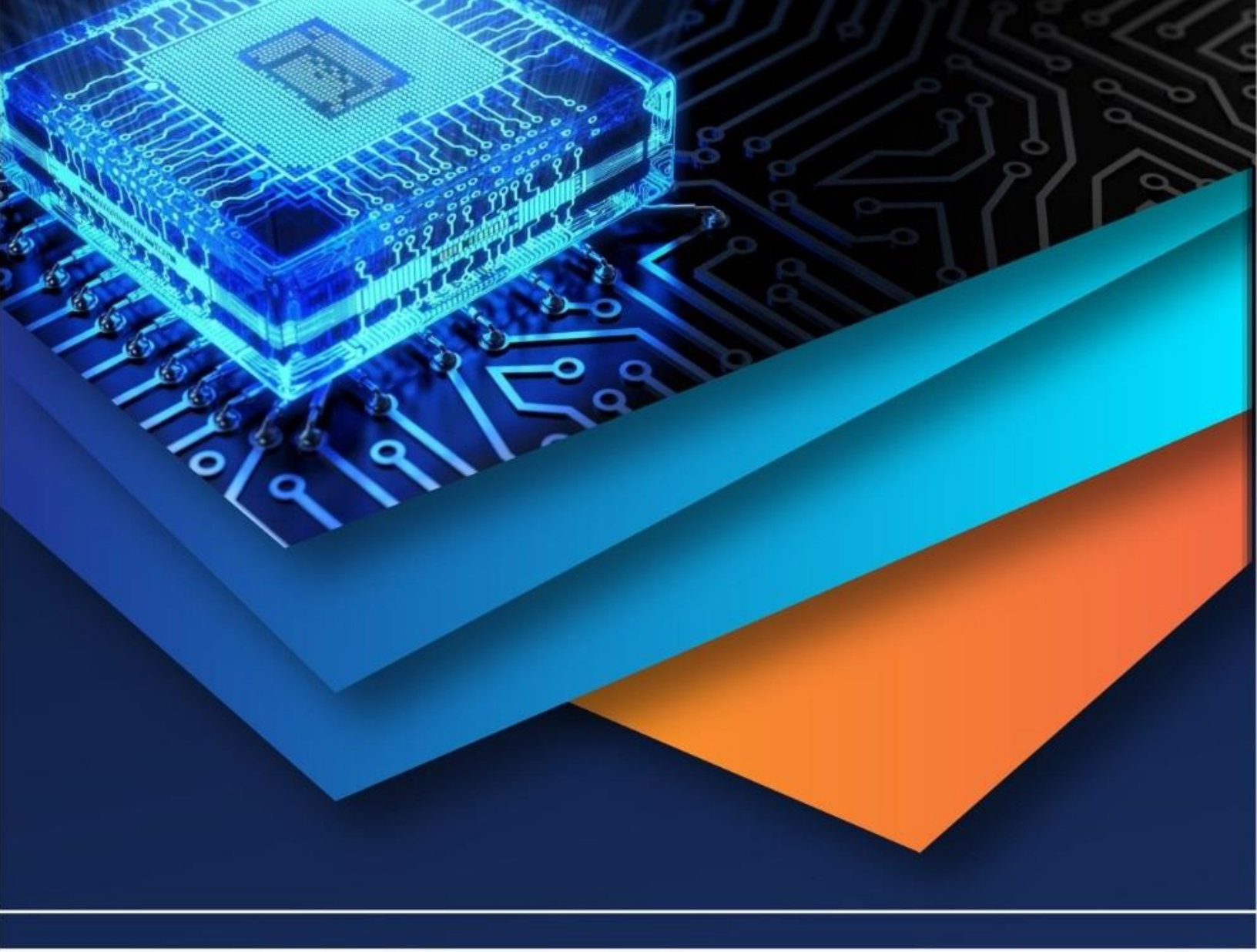
VIII. ACKNOWLEDGMENT

We express express their sincere gratitude to all those who contributed to the successful completion of this project, *"Fake News Detection Using BERT and RoBERTa,"* undertaken as part of the academic major project.The authors are especially thankful to their academic mentors and faculty members for their continuous guidance, technical insights, and encouragement throughout the project's development. Their expertise played a pivotal role in shaping the direction and success of this work.Appreciation is also extended to the open- source community for providing invaluable tools and frameworks such as Hugging Face Transformers, TensorFlow, PyTorch, and Flask, which significantly aided the model development, training, and deployment processes.

The authors would also like to acknowledge their peers and fellow students for their constructive feedback, collaborative discussions, and support during various stages of the project. Finally, heartfelt thanks go to the authors' families and friends for their constant encouragement, motivation, and support, which served as a source of strength throughout the project journey. This project represents a collective effort, and the authors are deeply grateful to everyone involved in making it a success.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, pp. 4171–4186, 2019.
[2] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
[3] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," in Proc. CIKM, pp. 797–806, 2017.
[4] S. Thorne et al., "FEVER: A Large-scale Dataset for Fact Extraction and Verification," Proc. NAACL-HLT, pp. 809–819, 2018.
[5] H. Zhang, P. Zhang, and Y. Yuan, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Network Model," IEEE T. Knowl. Data Eng., vol. 33, no. 5, pp. 2225– 2238, May 2021.
[6] A. Hanselowski et al., "Retrospective Fake News Detection Using BERT," Proc. NLP4IF@EMNLP, pp. 1–8, 2019.
[7] S. Karimi and J. Tang, "Deep Learning for Detecting Fake News in Social Media," IEEE Access, vol. 8, pp. 90594–90601, 2020.
[8] M. Zhou, W. Shu, D. Zhang, and J. Wu, "Fake News Detection via NLP Enhanced with Transformer-based Architectures," in Proc. IJCAI, pp. 4532–4538, 2021.
[9] D. Wadden et al., "Fact or Fiction: Verifying Scientific Claims," EMNLP, pp. 7534–7550, 2020.
[10] R. Shu, A. Sliva, H. Wang, and B. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," SIGKDD Explorations, vol. 19, no. 1, pp. 22–36, 2017.
[11] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A Deep Neural Network for Fake News Detection," IEEE Intelligent Systems, vol. 35, no. 4, pp. 45–50, 2020.
[12] S. Vaswani et al., "Attention Is All You Need," NeurIPS, pp. 5998–6008, 2017.
[13] A. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving Fake News Detection Using Entity Recognition and Emotion Classification," IEEE Access, vol. 8, pp. 100947– 100958, 2020.
[14] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.
[15] Z. Wang, C. Li, W. Zhang, and C. Cao, "Combining BERT with Knowledge Graph for Fake News Detection," IEEE Access, vol. 9, pp. 148514–148524, 2021.
[16] S. Jwa, H. Oh, K. Park, and M. Cha, "ExBAKE: Explainable Fake News Detection Using Knowledge- Enhanced BERT," in Proc. ACL, pp. 317–322, 2019.
[17] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," Proc. ACL, vol. 2, pp. 422– 426, 2017.
[18] Y. Zhou 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.
[19] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Dynamic Information for Fake News Research," Big Data, vol. 8, no. 3, pp. 171–188, 2020.
[20] N. A. Aslam, T. Nazir, and F. Saeed, "Fake News Detection using RoBERTa and Ensemble Learning," in Proc. ICAC, pp. 237–242, 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY