International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

Fake News Detection Using BERT

Anshu Aditya[1], B.V.S.S.Vardhan[2], D.S.Chanakya Varma[3], P.Kailashnadh Gupta[4], Dr Venkat Ramana M[5]

[1234]*Student, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.*
[5]*Assistant Professor, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.*

Abstract - *In our current global prone to changes and biased information environment, the most reliable and unbiased news is the means of rational decision-making and the world comprehension, and yet the increasing scale of fake news and partial reporting constitute a core challenge to the mass media credibility. We envisage creating a biased news article detector algorithm that will be powered by BERT, Google's pre-trained and powerful, natural language model. The strategy collects a diverse dataset of newsletters emanating from several sources relating to varied views on a vast group of topics with every composition classified as containing one of several biases like political bias, ideological bias, and sensationalism. The subsequent step boosts the performance of the pre-trained BERT model using this dataset, tweaking its parameters to deal with those thought-provoking features in the text data. Evaluating the trained model's performance through typical machine learning metrics like accuracy, precision, recall, and F1-score shows that it is indeed capable to effectively identify echoed-in biases in the text it is trained on, including way subtle hints, and an overall bias of the news. Therefore, this automatic system has the potential to help journalists, policymakers, as well as the general public have the right understanding regarding biased news media. Finally, the work focuses on creating state-of-the-art machine learning tools that can search and fix biased media content across all news texts by using BERT and advanced text analytics in order to check all the content for bias and promote transparency in media industry.*

***Key Words*: Machine learning, BERT model, News media credibility, Text analytics**

1.Introduction

The digital age has brought success stories in the search for information. Nowadays, social media networks and online news are seen as sources of information because change from traditional news is happening to many people. This gives everyone access to information and makes it very useful, but the environment that creates the media has become an environment for the spread of misinformation – “fake news”. Fake news, misinformation or disinformation presented as official news can take many forms: stories, packaged images, fake advertisements or fake news. This influence leads to the formation of pillars on social media known for decision making, social discourse, and religion. Global fake news problem: Fake news does not only cover countries and regions. In fact, space has the power to affect everyone, no matter where or who they are. However, special combinations are also available in some regions. undefined Major Internet Users: Because the Internet is limited and a significant portion of the population depends mostly on mobile devices for information, it is difficult for people to distinguish between right and wrong. .Linguistic Diversity: Powerful information storage operations at the heart of the country often fail to meet the needs of India's linguistic diversity. Political Polarization: In general, political schools like to use weapons, consider certain groups as "fake news", and then suppress the opposition with the "light problem" and "public opinion management" during elections. Social Trust Issues: As a result of some issues with media trust, people are easily influenced by fake news that they believe to be true, often due to their own stereotypes or biases. The Limits of the Law: A Project to Add to the Problem Currently, the way to combat fake news is mostly based on a single comparison and rules of thumb. These techniques often identify content or signatures frequently used by fake news organizations. Although it is useful in some situations, it also has limitations: Although it is useful in some situations, it also has limitations: Limited Adaptability: Solving this problem often requires clear solutions, solutions are not always made with new ones. Adoption of lie communication. As marketers get better at creating and identifying misinformation, content-based programs may not be as good at removing new information. Contextual blindness: Traditional methods often do not understand the integrity of the data. They may miss the difference between criticism and opinion, favor fake news, and lead to misclassifications. Language Barrier: Available solutions may not be able to speak local languages, slang and customs. They won't have time to track down fake news in a language other than English. Introduction to BERT: An excellent tool for handling the content of words BERT stands for Bidirectional Encoder Represented by transformers and is a family of pre learning, deep learning that demonstrates a good understanding of its meaning. Ability to use a word in a sentence. Unlike traditional methods, BERT teaches patterns

bidirectionally; This means that the content of a word depends on nearby words, including words to the left and right. This allows BERT to capture context and connections in text, making it a promising model for tasks that require good language skills, such as emotional analysis and writing. The Promise of BERT in News Distribution: About next steps and solutions. Understanding more details: BERT models are powerful designed to help determine the meaning of an article, especially whether a word is offensive or not. The meaning of the sentence or the whole meaning, Thoughts, feelings, etc. It allows them to separate real news from made-up stories, without any knowledge of the possibility of being involved in the preparation of the fake campaign. Adapting to changing strategies: BERT's model will grow as the strategies used by fake news continue to evolve. The skills they learn from lots of literature help them find used words or phrases that don't make the story seem new. Multi-language capability: BERT model uses single Language by default. By carefully considering different information about different languages, the same model can be modified to include the spread of fake news into different areas of conversation, ultimately finding a solution. Purpose The purpose of this project is to verify whether the BERT model is suitable for distinguishing real content from fake content on social media. We will create a BERT model that will provide training on information covering the fake news problem in our region. The performance of the model will be evaluated and analyzed in terms of its ability to explain the complexity of the selected words. By using BERT's artificial intelligence technology, we hope to create a more powerful and flexible system to combat fake news. This will lead to the emergence of a multicultural public opinion, which can be considered a new characteristic of our age.

2.Literature Survey

[1] Rahul Chauhan, Sachin Upadhyay and Himadri Vaidya titled “Fake News Detection based on machine learning algorithm”

Fake news has become a major problem in today's world, spread rapidly through social media. This misinformation can negatively impact public opinion and decision-making. To address this issue, researchers are exploring machine learning techniques for fake news detection. One approach involves analyzing the text of news articles using Natural Language Processing (NLP). This includes techniques like removing unnecessary words and converting the text into a format that machine learning algorithms can understand. Several machine learning algorithms have been studied for fake news detection. Some examples include Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Bidirectional Long Short-Term Memory (Bi-LSTMs). These algorithms can learn patterns from real and fake news data and then use those patterns to identify new fake news articles. The paper by Chauhan et al. proposes a system for fake news detection that utilizes a combination of machine learning algorithms. Their system involves collecting datasets of real and fake news articles, preprocessing the text data, converting the text into numerical vectors, and then applying machine learning algorithms to classify the news as real or fake. The study mentions using four algorithms in their model: Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. They achieved high accuracy (around 90-94%) in detecting fake news by combining the results from these algorithms. Logistic Regression provided the best individual accuracy (around 94%). While this approach shows promise, there are some limitations to consider. The paper doesn't specify the datasets used for training and testing the model. Additionally, it doesn't detail how the final outcome is determined by combining the results from the four algorithms. Overall, this paper highlights the potential of machine learning for tackling the problem of fake news. As research continues to develop, these techniques can become even more effective in helping us distinguish between real and fake news.

[2] Poonam Narang, Upasana Sharma titled “A Study on Artificial Intelligence Techniques for Fake News Detection”

Fake news is a growing problem on the internet, and its potential to harm society is significant. Researchers are actively developing methods to detect fake news, but this field is still in its early stages. This paper examines existing research on fake news detection techniques. The authors conducted a thorough analysis of various datasets used for fake news detection. They also explored the different techniques employed to identify fake news, including manual fact-checking by human experts and automated methods that leverage machine learning and artificial intelligence. These automated methods can analyze vast amounts of data, including text content, social network structures, and other relevant information. The review process involved examining over 200 research papers on fake news detection. From this collection, the authors selected 33 papers that focused on various detection techniques. Many of these techniques involve machine learning algorithms for classification, such as Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). Feature extraction is another crucial aspect, where researchers identify characteristics like language style, sentiment analysis, and topic modeling to differentiate real from fake news. Several publicly available datasets are used to train and test the effectiveness of these detection models. Some prominent examples include FakeNewsNet and LIAR.

The performance of these models is measured using metrics like accuracy, precision, recall, and F1 score. The paper also presents a comparative analysis of various state-of-the-art models. This analysis highlights the detection methods, datasets used, and performance achieved by different studies. However, the authors also identify several challenges that need to be addressed in future research. One challenge is the difficulty of accurately identifying the underlying social network structure, which limits the ability to predict how information spreads in the real world. Additionally, limited access to free and reliable API web services hinders the generation of trust factors for news sources. The rapid evolution of fake news on social media platforms necessitates faster detection techniques to stay ahead of this ever-changing threat. Extracting factual content from a mix of opinions and general statements also presents a significant challenge. Text normalization techniques might not be able to capture all temporal references, such as references to specific dates or times.

[3]Shubh Aggarwal, Siddhant Thapliyal, Mohammad Wazid, D. P. Singh titled "Design of a Robust Technique for Fake News Detection"

The vast amount of information available online makes it crucial to distinguish factual accuracy from misinformation. Truth detection models, powered by machine learning, are valuable tools for classifying statements as true or false. These models are used in various fields, including journalism, social media analysis, fact-checking, and legal investigations. Truth detection models offer several advantages. They automate the process of verifying information, saving time and resources. Additionally, they can handle the ever-increasing volume of data on the internet. These models also promote consistency by applying objective criteria for evaluating truthfulness, minimizing the influence of subjective judgments. Moreover, they complement human efforts by helping fact-checkers and investigators identify potentially false information. Researchers have actively explored various approaches for detecting fake news, including those based on content, social context, and existing knowledge. Some studies have focused on developing explainable decision systems for automated fake news detection, while others have addressed challenges related to imbalanced data in training models. Despite their advantages, truth detection models have limitations. The accuracy of these models can be affected by the quality and availability of training data. Additionally, there's a need for further research to improve how these models explain their reasoning behind classifications (interpretability). Furthermore, as tactics for spreading misinformation evolve, new techniques are needed to stay ahead of these ever-changing challenges. This survey provides a comprehensive overview of truth detection models, highlighting their applications, limitations, and potential areas for future research. It serves as a foundation for understanding the current state of the art and paves the way for further exploration in this critical field

[4]Yadong Gu, Mijit Ablimit, Askar Hamdulla titled "Fake News Detection based on Cross – Model Co - Attention"

Rumors spread quickly online, and with the constant development of social media, the format of these rumors has evolved. They are no longer just text-based but often combine text and images to create a more convincing facade. This makes it crucial to have effective detection methods. Traditionally, rumor detection relied on analyzing textual features and employed machine learning algorithms for classification. However, with the rise of deep learning, neural networks have become the go-to approach for feature extraction and classification in rumor detection tasks. These models can capture various aspects of textual data, including temporal information, structure, and linguistic cues. Recurrent neural networks (RNNs) are particularly useful for learning hidden representations from sequential text data like tweets, while convolutional neural networks (CNNs) can identify key features scattered within the text. Despite the advancements in text-based rumor detection, there's a limitation: relying solely on text might not be sufficient for accurate judgment. Fake news often leverages the combined power of text and images on social media platforms. This has led to the rise of multimodal rumor detection, which recognizes the importance of combining textual and visual information for better detection accuracy. Multimodal rumor detection explores different techniques for fusing these two modalities. Early fusion combines features from text and image before feeding them into the model, while late fusion combines features after processing them separately. Attention mechanisms are also being explored to focus on the most relevant aspects within text and image features. However, there are still challenges to overcome. Existing models might not fully capture the intricate relationship between text and image content. More research is needed on methods that can effectively exploit the interaction between these modalities. Additionally, extracting textual information embedded within images itself could be a valuable avenue for future exploration in multimodal rumor detection. This survey provides a comprehensive overview of the evolution of rumor detection models, highlighting the limitations of unimodal approaches and the potential of multimodal methods for more accurate detection of fake news.

3. Analysis and Design

The proliferation of fake news online poses a significant threat to public discourse and informed decision-making. This survey explores the potential of combining Bidirectional Encoder Representations from Transformers (BERT) for improved detection. Detecting the veracity of online information is complex due to factors like fabricated content, emotional manipulation, and rapid dissemination. Existing approaches include machine learning algorithms like naive Bayes classifier, logistic regression, and support vector machines, alongside natural language processing techniques. BERT, a powerful language model, excels at understanding text nuances, making it suitable for analyzing news articles and identifying potential falsehoods. Studies have shown positive results in using BERT for tasks like sentiment analysis and text classification which requires understanding of the language. Limitations and Challenges include bias in training data and models, necessitating fairness to avoid inaccurate results. The evolving nature of fake news tactics requires continuous adaptation of detection methods. Research on combining BERT with other AI models holds promise for further improvement, along with developing explainable AI approaches to understand model conclusions and address biases. Addressing bias in both data and models is essential for fair and responsible development and deployment of fake news detection systems. Combining BERT shows potential for improved fake news detection. However, addressing limitations like bias and the evolving nature of fake news tactics is crucial for responsible development and deployment of such technologies. Exploring additional AI models and explainable AI approaches can further contribute to advancing this field.

3.1 Terminologies

3.1.1 Logits

These are one of the most common, especially used during proportions. Logits are estimates that are then normalized using the softmax distribution. In a classification problem, the model associates the input with the probability for each class. The logit term represents the probability of the prior model before applying the softmax function designed to convert the raw score into probability. "Logit" is derived from the logistic function, a regression model commonly used in binary problems. For most register logic, it usually contains a score vector where each score is specific to a class. These raw scores are then converted into results with the help of the softmax function to ensure a consistent result during the application.

3.1.2 Sigmoid Function:

A sigmoid function is a bounded, differentiable, real function that is defined for all real input values and has a non-negative derivative at each point. It has a characteristic S-shaped curve or sigmoid curve. This function takes any real-valued number and "squashes" it into a value between 0 and 1. This is particularly useful when we want to interpret the output of our model as a probability. The sigmoid function is monotonic and has a first derivative, which is bell-shaped. It has exactly one inflexion point.

3.2 BERT

BERT (short for Bidirectional Encoder Representation called Transformers) is an ML (machine learning) model for natural language processing. Developed by Google AI researchers in 2018, BERT is a multi-purpose solution that solves mor e than 11 of the most commonly used tasks, such as responsibility-related sentiment analysis and recognition. Traditionally computers have been great at collecting, storing and reading data, but they have faced language understanding problems. Improve natural language processing (NLP) skills Smart computers use natural language processing (NLP) technology to read and understand spoken and written language. This integrated method converts the knowledge of words, numbers, and procedures used by computers into the syntax of the human language. In general, an NLP task is solved only with a model designed for a specific purpose. But BERT successfully solves more than 11 NLP tasks, changing the NLP state, surpassing their performance and becoming versatile and adaptable to many languages. The model is designed for deep, bidirectional representation of content. Therefore, computer scientists can add an output layer to BERT to create global models for various NLP tasks.

4.Dataset Description

The Indian Fake News Detection (IFND) Dataset: A Resource for Political News Classification This project utilizes the Indian Fake News Detection (IFND) dataset, a collection of real-world news articles pertaining specifically to India. The dataset leverages content scraped from reputable Indian fact-checking websites. It comprises two distinct categories: real news and fake news articles.

The IFND dataset offers a valuable resource for training and evaluating machine learning models focused on political news classification within the Indian context. The articles cover a variety of topics, with a concentration on political news, reflecting the prevalent nature of such content in the Indian news landscape.

S.No	Attribute	Description
1.	ID	Unique identifier for each news
2.	Statement	Title of the news article
3.	Image	Image Url
4.	Category	Topic of news
5.	Date	Date of news
6.	Label	1 indicate True news and 0 indicate fake news

Table 1: Dataset attributes and Description

5.Tools And Libraries

5.1 Seaborn

Seaborn, based on matplotlib and used as a powerful data visualization package, is a library which is developed for the same purpose. It provides a tall-level interface through which one may be intact as well as provide Impressive statistical graphics.

5.2 Scikit-learn

Within the Python data science community, Scikit-learn is the most widely utilized machine learning library. It has a reputation for its user-friendly interface, compatibility with other well-known scientific Python libraries like NumPy and Matplotlib, and a broad range of tools for data analysis positions. For many machine learning enthusiasts and data scientists, scikit-learn is their preferred solution because of these features.

5.2 Pandas

Pandas is a free library for data management and analysis using the Python language. It provides two data elements: List (one-dimensional inline array) and DataFrame (two-dimensional inline data structure consisting of rows and columns).

5.3 NumPy

NumPy (Numerical Python) is an important module for computing in Python. It has built-in support for large multidimensional arrays and matrices, as well as a number of advanced mathematical functions for manipulating arrays.

5.4 Transformers

Transformers library is a deep learning NLP library developed by Hugging Face using the best of natural language processing (NLP). Text classification, language generation, name recognition, responsiveness, etc. It is a repository of pre-learning models and tools suitable for most NLP tasks, such as Some key features and functions of the Transformers library are: Some key features and functions of the Transformers library are: Training opportunities in many languages, including BERT, GPT-2, RoBERTa, XLNet and others, to name just a few. This process of learning big data can be modified to work or record, depending on its limitations.

5.5 Matplotlib

Matplotlib is a useful tool for visualizing data (including charts and graphs) in Python. It provides a low-level API that allows the creation of many different static, graphical and interactive functions in Python.

6.Working Of Model

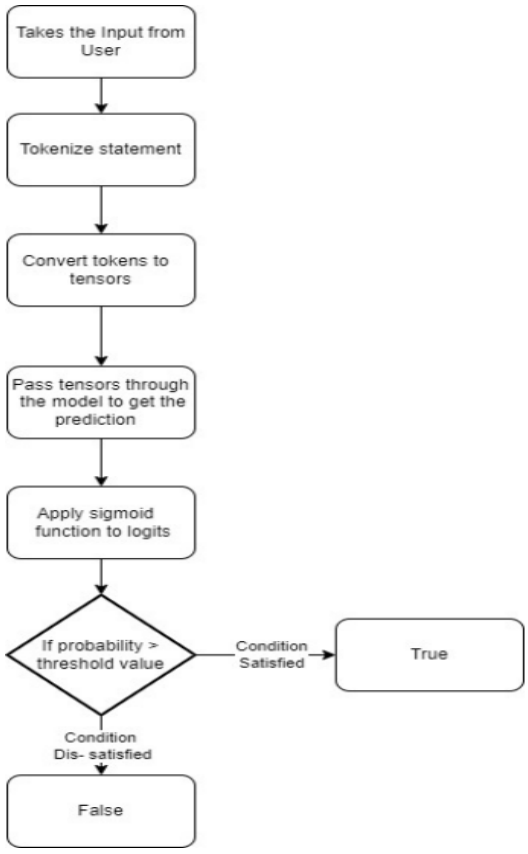


Figure 1: Model Flowchart

When the program starts the user will give an input statement and the statement is passed on to the model. The model will tokenize the statement into words or sub-words. These tokens are then converted into Tensors, which are multi-dimensional vectors. These Tensors are then passed into the model, which must predict the probability. To do this, the model converts the Tensors into Logits, commonly used in machine learning, particularly in classification tasks. The logits will then be given to the Sigmoid function which will give the output as probabilistic value. After setting a Threshold Value the model will compare it with the probability and return the output.

7.Result

Upon successful training, we tested our model with some common statements which are not present in the dataset, and we got some positive results. Model Performance on Unseen Data To assess the model's generalization capabilities, we tested it on statements not present in the training dataset. We included examples like:

Statement 1: MS Dhoni was the captain of Indian cricket team

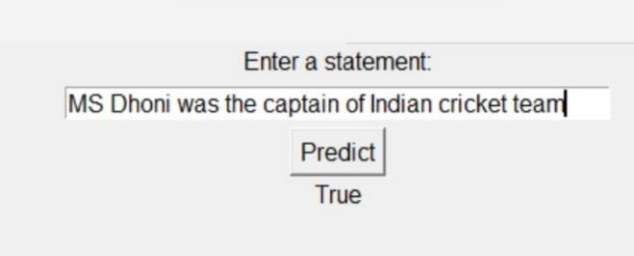


Figure 2: Result for statement 1

Statement 2: WhatsApp news is reliable

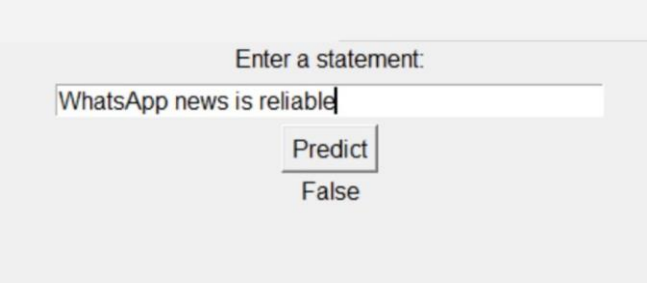


Figure 3: Result for statement 2

These statements highlight the model's ability to go beyond simple keyword matching and leverage its understanding of context and language relationships for classification.

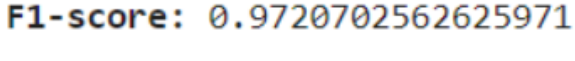
7.1 Model Evaluation:

To comprehensively evaluate the model's performance, we measured key metrics:

Accuracy: This metric reflects the overall percentage of correct predictions made by the model.

Precision: This metric measures the proportion of positive predictions that were actually correct (true positives / (true positives + false positives)). It indicates how good the model is at avoiding false positives (predicting real news when it's fake).

F1 Score: This metric combines precision and recall (the proportion of actual positive cases the model identified correctly) into a single measure, providing a balanced view of the model's performance.



Confusion Matrix

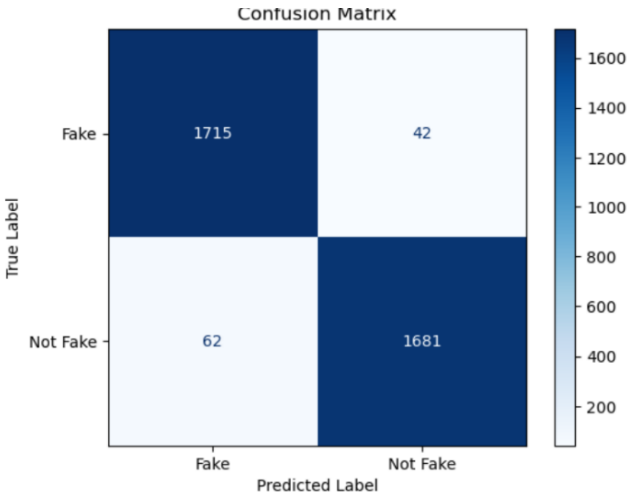


Figure 4: Confusion Matrix:

The confusion matrix shows the performance of our model classifying news articles as real or fake. The text labels on the axes indicate the Fake or Not Fake(True) labels. The values in the table represent the number of news articles that fall into each category.

Breakdown of the values in the matrix:

True Positives (TP): 1715 - The model correctly classified 1715 real news articles as real.

False Negatives (FN): 62 - The model incorrectly classified 62 real news articles as fake.

False Positives (FP): 42 - The model incorrectly classified 42 fake news articles as real.

True Negatives (TN): 1681 - The model correctly classified 1681 fake news articles as fake.

Classification Report:

	precision	recall	f1-score	support
0.0	0.97	0.98	0.97	1757
1.0	0.98	0.97	0.97	1743
accuracy			0.97	3500
macro avg	0.97	0.97	0.97	3500
weighted avg	0.97	0.97	0.97	3500

Figure 5: Classification Report:

8. Limitations

Dataset Scope: Our model is currently trained on a dataset focused on Indian news and headlines. This limits its knowledge base to news content specific to that region. To broaden its applicability, future work could involve incorporating data from various regions and languages.

Computational Resources: Fine-tuning BERT on even larger datasets requires significant computational resources, including powerful GPUs and faster memory. Exploring techniques for efficient model training and optimization will be crucial for future improvements.

9. Conclusion

A Fake News Detection system was implemented using BERT to show how state-of-the-art natural language processing models can be used so that the task can be performed more efficiently. The project began with preprocessing our classified dataset; attention, however, was given to addressing any biases in terms of class distribution and also checking on data integrity. The neural network is designed for binary classification, using the powerful BERT(Bidirectional Encoder Representations from Transformers) model in order to differentiate between fake and genuine statements. This enhanced the performance of the model due to its ability of capturing contextual information and semantic nuances.

During training, there were several hyperparameters being tuned carefully for instance by applying dropout as a method of regularization and optimizing the model through Adam optimizer. It's critical during the training loop process that these parameters should be iteratively adjusted so that Binary Cross-Entropy loss is minimized, this way making sure that input's true meaning is learnt by the model. Evaluation stage provided metrics such as precision, recall, F1-score through a comprehensive classification report that showed how well the model could generalize on unseen data. Consequently, quality results clearly indicate that BERT-based approach is highly effective in identifying fake news utterances

10.Future Work

Integration with APIs and Bias DetectionTo enhance the model's longevity and adaptability

API Integration: We can integrate the model with news APIs to continuously expose it to fresh data and news streams. This approach helps the model stay relevant and adapt to evolving trends in news content.

Bias Detection: By training the model to identify potential biases within news articles retrieved from APIs, we can provide a more nuanced analysis of the information. This can empower users to make informed judgments about the credibility of news sources.

11.Reference

[1] Rahul Chauhan, Sachin Upadhyay and Himadri Vaidya titled "Fake News Detection based on machine learning algorithm"

[2] Poonam Narang, Upasana Sharma titled "A Study on Artificial Intelligence Techniques for Fake News Detection"

[3] Shubh Aggarwal, Siddhant Thapliyal, Mohammad Wazid, D. P. Singh titled "Design of a Robust Technique for Fake News Detection"

[4] Yadong Gu, Mijit Ablimit, Askar Hamdulla titled "Fake News Detection based on Cross – Model Co - Attention"