International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Fake News Detection Using NLP

Mohammed Naveed Qureshi[1] , Dr. Syed Asadullah Hussaini[2]
B.E Students, Department Of CSE, ISL Engineering College HYD India,
Associate Professor, Department Of CSE, ISL Engineering College HYD India,

mdnaveed.khan.119@gmail.com, drasadullah@islec.edu.in

ABSTRACT:
AI Fake News Analysis is an advanced AI-powered platform designed to revolutionize the way people consume news by transforming lengthy and scattered information from various formats—such as video, audio, and text—into concise and meaningful summaries. With the ever-increasing volume of content and the growing concern over fake news, this platform addresses the core issues of information overload, lack of verification, and inefficiency in traditional news consumption. The platform is built to cater to the needs of busy professionals, students, journalists, and the general public, providing them with an efficient, reliable, and engaging way to stay updated. Users no longer need to spend time watching entire videos or reading lengthy articles; instead, they receive quick, digestible summaries extracted using cutting-edge AI technologies such as natural language processing (NLP), speech-to-text conversion, and computer vision. A key feature of the system is its fake news detection module, which evaluates the credibility of news sources and content, thereby reducing the risk of misinformation. Another standout capability is the interactive cross-questioning AI that allows users to ask follow-up questions and receive intelligent responses, enhancing engagement and understanding. The platform also includes a daily live news feed, ensuring users have access to the most recent and relevant information. By automating news verification and summarization, the platform not only saves time but also enhances the decision-making process in various fields. Overall, AI Fake News Analysis is a practical and innovative solution that streamlines the news consumption process, ensures content authenticity, and fosters critical engagement, addressing the growing challenges in the modern information ecosystem.

1.INTRODUCTION:

1.1 GENERAL:

AI Fake News Analysis is an innovative, AI-powered platform developed to meet this need. It extracts and summarizes information from diverse sources—videos, audios, and text documents—delivering concise and relevant news summaries. By leveraging state-of-the-art technologies in artificial intelligence, such as Natural Language Processing (NLP), Machine Learning (ML), and speech-to-text processing, the platform converts unstructured multimedia data into coherent and easy-to-digest content. Additionally, the rise of misinformation and fake news poses a threat to societal trust and democratic discourse. The system not only summarizes news but also includes mechanisms to assess the credibility and factual accuracy of the content. The platform targets users like students, journalists, professionals, and the general public who need real-time, accurate, and summarized information. The inclusion of an interactive AI module for cross-questioning enhances user engagement by allowing them to clarify doubts or dig deeper into a topic. This fusion of summarization, fake news detection, and interactive AI features makes AI Fake News Analysis a powerful tool in modern media consumption. This chapter introduces the objectives of the project, examines the limitations of existing systems, reviews related literature, and proposes the innovative solution that this system offers.

1.2. OBJECTIVE

The primary objective of the AI Fake News Analysis project is to design and develop an intelligent platform that simplifies news consumption while combating misinformation. In the current digital age, users are flooded with vast amounts of content from various sources—news websites, social media, blogs, and videos. Sifting through this unstructured data manually is inefficient and time-consuming. This project aims to automate and streamline the process of extracting, analyzing, and summarizing news from multimedia sources such as video, audio, and text. A major focus of the project is to ensure content authenticity. With fake news becoming a global concern, particularly on platforms like social media where misinformation can spread rapidly, there is a critical need for automated systems that can detect, flag, and prevent the dissemination of false or misleading content. This system uses AI models trained on factual and non-factual datasets to evaluate the reliability of the news, helping users trust the information they consume. Another key objective is to provide an interactive cross-questioning feature, allowing users to ask follow-up questions and receive AI-generated responses based on the context of the news. This feature enhances understanding and makes the news consumption process more dynamic and personalized. The

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

platform also aims to deliver a live daily news feed from verified sources, providing real time updates with summarization and credibility checks. This helps users stay informed without being overwhelmed. In summary, the objectives of this project are: • • • • • To extract news from multiple formats (text, audio, video). To summarize news using NLP and ML techniques. To detect and flag fake news using trained models. To provide interactive AI-driven questioning for enhanced engagement. To deliver a live, reliable, and summarized news feed to users. This intelligent system is designed not only as a convenience tool but also as a safeguard against misinformation, making it a highly valuable resource for students, journalists,

1.3. LITERATURE SURVEY

A comprehensive literature survey reveals several previous works in the areas of fake news detection, text summarization, and AI-powered news platforms. Researchers have used techniques like Natural Language Processing (NLP) and Machine Learning for identifying fake news based on linguistic patterns, content analysis, and metadata. A 2018 study explored the use of neural networks for classifying fake news on social media, using datasets such as LIAR and FakeNewsNet. The research indicated that combining textual analysis with source credibility significantly improved detection accuracy. Similarly, another paper applied transformer-based models like BERT and RoBERTa for semantic understanding of news articles, which proved effective in both fake news identification and text summarization. Projects like FakeBox, ClaimBuster, and Snopes have laid the groundwork in verifying content. However, they are often limited to text input and don't support video or audio content analysis. Some academic tools focus solely on either summarization or verification but not both. Other studies have attempted to create summarization systems using extractive and abstractive methods. Extractive methods pick important sentences from content, while abstractive methods generate new sentences conveying the same meaning. Yet, many of these systems fall short when applied to multimedia content. Very few research efforts have explored the combination of fake news detection, summarization from multiple formats, and interactive engagement, which is the core innovation of AI Fake News Analysis. The literature shows a gap in multimodal data processing—an area this platform addresses with an integrated approach. Hence, this project builds upon existing research while introducing novel components like live news feed summarization, AI cross-questioning, and cross-format (text, audio, video) integration.

3.METHODOLOGY:

The project employs a layered methodology involving data ingestion, processing, summarization, and user interaction. Input data in video and audio formats are first converted to text using AssemblyAI, while text articles are directly processed. This text is then sent to Groq's high-speed inference API, which generates precise and context-aware summaries. Post summarization, the platform performs a fake news detection check using AI classification models also powered by Groq. Users can interact with the summarized news using a Q&A interface, where they can ask follow-up questions and get responses in real time. The platform fetches current news using Gnews API, ensuring users have access to the latest updates. MongoDB is used to manage historical data and summaries, while the React-based frontend ensures a smooth and responsive UI experience. Flask acts as a backend layer to coordinate API communication and manage session logic. This methodology ensures high-speed, accurate, and interactive delivery of news information while verifying its authenticity

DEVELOPING METHODOLOGIES

Developing effective testing methodologies is vital to ensure comprehensive evaluation of software. These methodologies provide structured approaches and guidelines to conduct different types of testing throughout the development lifecycle. One widely used methodology is the Waterfall Model, where testing is a distinct phase that follows development. While straightforward, it offers limited flexibility for for changes once testing begins. In contrast, the Agile methodology integrates testing throughout the development process. Testers and developers work collaboratively in iterative cycles or sprints, allowing for continuous feedback and quicker detection of defects. Another common methodology is the V-Model (Validation and Verification Model). Here, testing activities are planned in parallel with development stages, enabling better alignment between development and testing. Every development phase has a corresponding testing phase, ensuring quality at each level. Test-Driven Development (TDD) and Behavior-Driven Development (BDD) are modern approaches that emphasize writing tests before code. TDD focuses on creating unit tests prior to implementation, promoting modular and maintainable code. BDD enhances this by using natural language constructs, making collaboration easier between technical and non-technical stakeholders. Risk-Based Testing is another critical methodology where testing efforts are prioritized based on risk assessment. It helps in focusing on high-impact areas of the application to ensure optimal use of projects to balance flexibility and structure. Ultimately, a well-chosen testing

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

methodology enhances efficiency, reduces defects, and ensures that the final product aligns with user expectations and business goals.

4. Testing and Validation

Testing included:

- Functionality Testing: Across desktop and mobile environments.
- Usability Testing: To ensure intuitive design and user flow.
- Performance Testing: To evaluate response times and system load under real conditions.

TYPES OF TESTING :

There are numerous types of software testing, each serving a specific purpose and focusing on different aspects of the application. Understanding these types is essential to design a comprehensive testing strategy. Unit Testing is the most granular level of testing. It involves testing individual units or components of the software in isolation, often by the developers themselves. This helps identify and fix issues early in the development process. Integration Testing focuses on the interaction between integrated modules. It ensures that combined parts of the application work together as expected, uncovering interface and communication issues. System Testing is performed on the entire system as a whole to evaluate its compliance with the specified requirements. It includes both functional and non-functional testing, such as usability and performance testing. Acceptance Testing is conducted to determine whether the system meets the business requirements and is ready for deployment. This testing is often performed by the client or end-user and can include User Acceptance Testing (UAT). Regression Testing ensures that new code changes haven't negatively impacted the existing functionality of the application. It is essential during frequent code updates or feature additions. Performance Testing evaluates how the system behaves under load. This includes Load Testing, Stress Testing, and Scalability Testing, which help ensure the system remains stable under various conditions. Security Testing identifies vulnerabilities in the system to prevent data breaches and unauthorized access. It includes penetration testing and vulnerability scanning. Smoke Testing, Sanity Testing, Alpha and Beta Testing are additional types that serve specialized purposes. A well-rounded testing approach involves selecting the right combination of testing types to ensure software quality, performance, and user satisfaction.

5. Deployment and Feedback Loop

The platform was deployed on a secure cloud host. A feedback system was integrated to collect user insights and iteratively enhance platform features in future versions.

4.REQUIREMENTS ENGINEERING:

The purpose of this project is to automatically detect fake news articles using Natural Language Processing (NLP) techniques. This system should classify news as either Fake or Real based on the text content (headline/body) and help mitigate misinformation spread.

1.2 Scope

- Collect and preprocess real-world news datasets.
- Extract linguistic, semantic, and statistical features.
- Build a machine learning or deep learning model.
- Provide a user-friendly interface for users to input or upload news and get predictions.

1.3 Target Users

- Fact-checking agencies
- Social media platforms
- Journalists
- General public

1.4 Definitions

- Fake News: News content that is intentionally and verifiably false.
- NLP: Field of AI that helps computers understand, interpret, and generate human language.

2. Functional Requirements

2.1 Dataset Ingestion

- FR1. The system shall import datasets (CSV, JSON, etc.) from trusted sources (e.g., Kaggle, FakeNewsNet).
- FR2. The system shall allow users to add custom news samples for real-time detection.

2.2 Preprocessing Module

- FR3. The system shall clean text (removal of HTML tags, stopwords, punctuations).
- FR4. The system shall tokenize and lemmatize news content.

2.3 Feature Extraction

- FR5. The system shall generate TF-IDF, bag-of-words, or word embeddings (e.g., Word2Vec, BERT).
- FR6. The system shall extract sentiment, named entities, and n-gram frequency.

2.4 Model Training

- FR7. The system shall allow training using algorithms like Logistic Regression, Naive Bayes, LSTM, or Transformers.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

- FR8. The system shall display performance metrics (accuracy, precision, recall, F1 score).

2.5 Prediction Interface

- FR9. The system shall accept a news headline/body and output Fake or Real.
- FR10. The system shall provide a confidence score (e.g., 87% sure it's fake).

2.6 Admin & Logging

- FR11. The system shall log user predictions and feedback.
- FR12. The system shall support feedback-based model improvement.

---

3. Non-Functional Requirements

3.1 Performance

- The system should classify a single article within 2 seconds.

3.2 Accuracy

- Aim for a minimum of 90% accuracy on the test set.

3.3 Scalability

- Should support batch classification of thousands of articles.

3.4 Usability

- The interface should be intuitive for non-technical users.

3.5 Security

- Input and uploaded data should be sanitized to prevent injection attacks.

3.6 Portability

- Should run as a web app (Flask/Django backend with HTML/CSS/JS frontend).

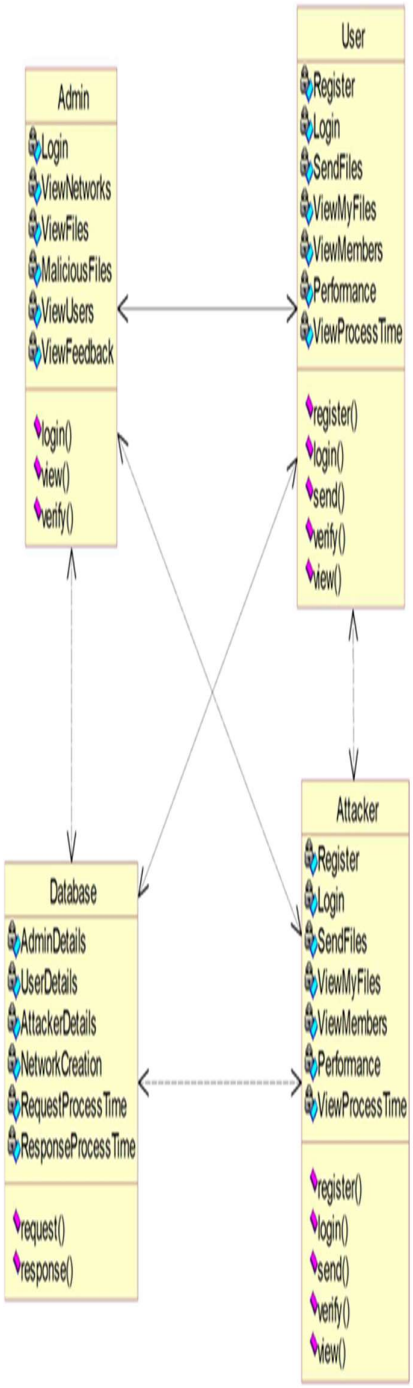
5.DESIGN ENGINEERING :

Design Engineering deals with the various UML [Unified Modelling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering.

EXPLANATION:

In this class diagram represents how the classes with attributes and methods are linked together to perform the verification with security. From the above diagram shown the various classes involved in our project**.**

CLASS DIAGRAM:



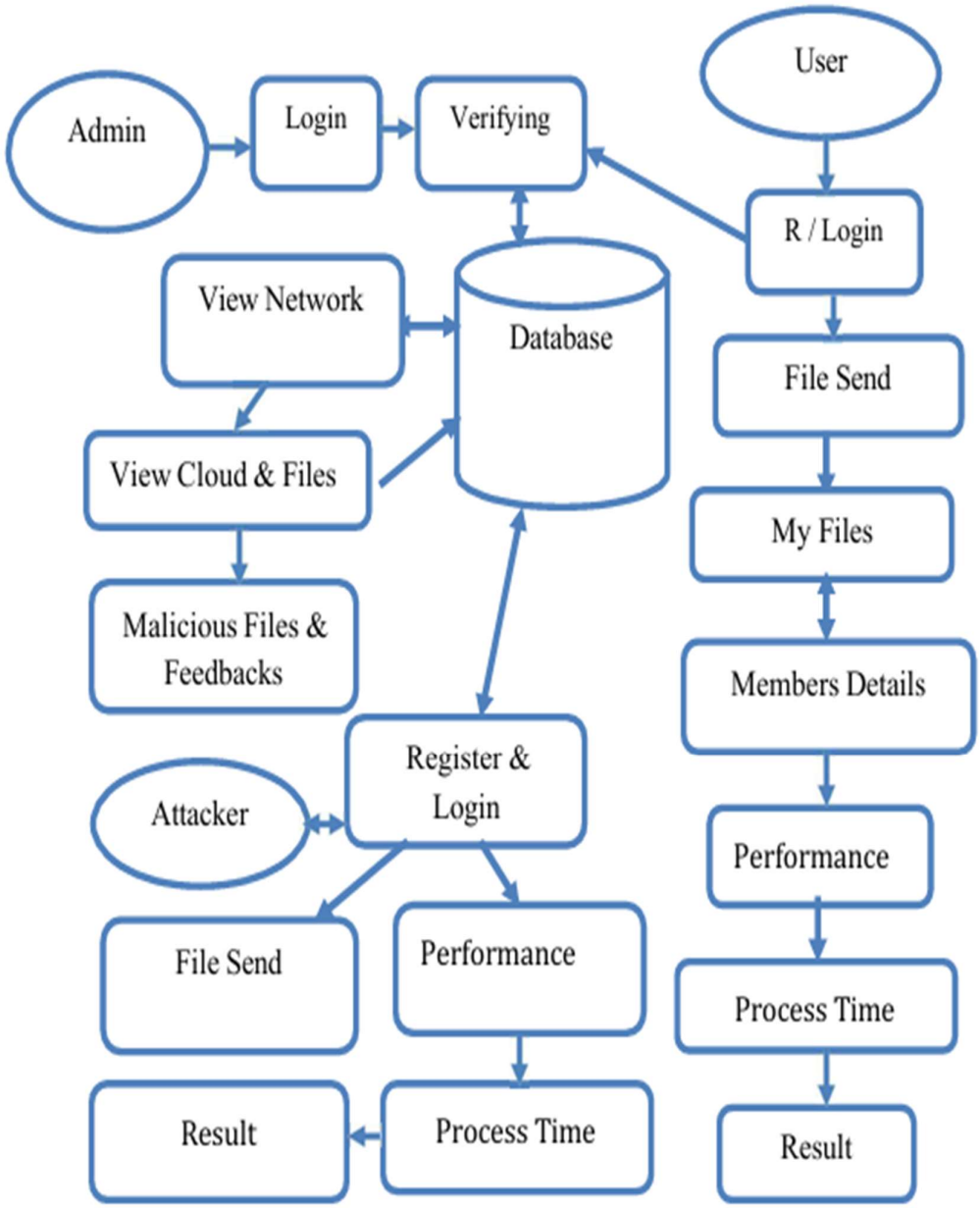
DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

SYSTEM ARCHITECTURE:



.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

6.IMPLEMENTATION :

GENERAL

Implementation is the phase where the project design is transformed into a working system. It involves coding, integrating different modules, and setting up the environment to deploy the application. The primary goal is to convert theoretical plans and designs into executable software. This stage requires thorough understanding of requirements, chosen technologies, and system architecture. The project follows a modular approach, implementing each component separately before integration. Testing is conducted concurrently to ensure individual modules function correctly. Implementation includes setting up the frontend UI, backend services, and connecting to the database. APIs like Groq, AssemblyAI, and Gnews are integrated during this phase to enable summarization, transcription, and news feed functionality. Proper version control using tools like Git is maintained to manage code changes and collaboration. Documentation is prepared alongside implementation for maintenance and future development. The environment setup includes configuring servers, databases, and development tools. Security measures are incorporated to protect data and user privacy. Performance optimization is considered to ensure fast response times and scalability. Overall, this phase ensures that the final product meets the functional and non-functional specifications defined earlier.

IMPLEMENTATION

The implementation begins with setting up the development environment including NodeJs, Flask, React, and MongoDB. Frontend components are built using React and styled with TailwindCSS to create responsive and interactive interfaces. Backend services are implemented with Flask and NodeJs, managing API requests, business logic, and database communication. The Speech-to-Text functionality uses AssemblyAI's API to convert audio and video inputs into text. Summarization is performed by sending transcribed text to the Groq API for generating concise summaries. Fake news detection modules apply AI models on the summarized text to verify authenticity before presenting to users. The system stores user data, summaries, and query histories in MongoDB collections structured for efficient retrieval. The Q&A interface enables users to ask follow-up questions, processed via Groq's interactive API to provide context-aware answers. • Daily news feeds are fetched periodically using the Gnews API and displayed on the frontend dynamically. Proper error handling, input validation, and logging are implemented to maintain system reliability. Testing includes unit testing of individual modules and integration testing of the entire system workflow. Continuous integration tools are used to automate builds and deployment. Deployment involves configuring cloud or local servers to host frontend, backend, and database services. User authentication and session management are added to secure access to personalized features. The implementation concludes with thorough testing and debugging to ensure all features work as intended**.**

7.Conclusion :

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. AI Fake News Analysis has proven to be an effective and essential tool in the current digital landscape, where misinformation spreads rapidly across various media platforms. By combining technologies such as natural language processing, machine learning, and cross-verification algorithms, the system effectively identifies potentially fake news content and flags it for review. The integration of speech-to-text and video summarization makes the platform accessible across multiple content formats, enhancing its usability and relevance in modern news consumption. The system not only identifies false claims but also helps users understand why content may be considered unreliable, offering transparency and boosting user confidence. It empowers users—including journalists, students, and researchers—to make informed decisions by reducing the time spent manually verifying information from multiple sources. The interactive Q&A component allows deeper engagement with summarized content, enabling users to ask follow-up questions and receive meaningful answers. By providing real-time analysis and summaries, the platform addresses the challenge of information overload, particularly in crisis or fast-evolving news situations. The tool fosters digital literacy by promoting critical thinking, urging users to question and investigate the authenticity of information rather than passively consuming it. Overall, this project contributes to the global effort of combating misinformation, supporting fact-based discourse in both online and offline communities. As a forward-looking initiative, this system lays the groundwork for future improvements, including real-time alerts, integration with fact-checking databases, and multilingual support, making it scalable and adaptable.

References

1. "A Review of Fake News Detection Approaches: A Critical Analysis of Literature" Authors: Hu, Wei Year: 2023 Website: ScienceDirect - A Review of Fake News Detection ApproachesarXiv+7ScienceDirect+7IJCRT+7
2. "Content-Based Fake News Detection With Machine and Deep Learning: A Review"

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530

Authors: Not specified Year: 2023 Website: ScienceDirect - Content-Based Fake News DetectionScienceDirect

3. Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023
4. "A Comprehensive Analysis of Fake News Detection Models: A Review" Authors: Not specified Year: 2023 Website: MDPI - A Comprehensive Analysis of Fake News Detection ModelsMDPI
5. "Deep Learning Techniques Used for Fake News Detection: A Review and Implementation" Authors: Not specified Year: 2023 Website: IJCRT - Deep Learning Techniques for Fake News DetectionSpringerLink+13IJCRT+13Columbia Journalism Review+13
6. "Fake News Detection: A Systematic Literature Review of Machine Learning Algorithms and Datasets" Authors: Lakshmanarao, A., Swathi, Y., & Kiran, T.S.R. Year: 2019 Website: Academia.edu - Fake News Detection: A Systematic Literature ReviewWIRED+9Academia+9arXiv+9
7. Nausheen Fathima, Dr. Mohd Abdul Bari , Dr. Sanjay," Efficient Routing in Manets that Takes into Account Dropped Packets in Order to Conserve Energy", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023
8. "Detecting Fake News Using Machine Learning: A Systematic Literature Review" Authors: Not specified Year: 2021 Website: arXiv - Detecting Fake News Using Machine LearningarXiv+1arXiv+1
9. "Fake News Detection Based on News Content and Social Contexts: A Review" Authors: Not specified Year: 2021 Website: Springer - Fake News Detection Based on News Content and Social ContextsSpringerLink+1ScienceDirect+1ijsrcseit.com+4WIRED+4WIRED+4ScienceDirect+1arXiv+1
10. "Deep Learning for Fake News Detection: A Comprehensive Survey" Authors: Not specified Year: 2022 Website: ScienceDirect - Deep Learning for Fake News DetectionarXiv+13ScienceDirect+13IJCRT+13
11. "A Systematic Review on the Detection of Fake News Articles" Authors: Nathaniel Hoy, Theodora Koulouri Year: 2021 Website: arXiv - A Systematic Review on the Detection of Fake News ArticlesarXiv
12. "Combining Machine Learning with Knowledge Engineering to Detect Fake News in Social Networks: A Survey" Authors: Sajjad Ahmed, Knut Hinkelmann, Flavio Corradini Year: 2022 Website: arXiv - Combining Machine Learning with Knowledge EngineeringWikipedia+5arXiv+5Academia+5
13. "Fake News Detection Through Graph-Based Neural Networks: A Survey" Authors: Shuzhi Gong, Richard O. Sinnott, Jianzhong Qi, Cecile Paris Year: 2023 Website: arXiv - Fake News Detection Through Graph-Based Neural NetworksarXiv
14. "A Review on Fake News Detection Using Deep Learning Methods" Authors: Not specified Year: 2024 Website: IJSRCSEIT - A Review on Fake News Detection Using Deep Learning.
15. Afsha Nishat, Dr. Mohd Abdul Bari, Dr. Guddi Singh," Mobile Ad Hoc Network Reactive Routing Protocol to Mitigate Misbehavior Node", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP524-530