Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

NARADMUNI: AN AI-BASED INTELLIGENT FAKE NEWS DETECTION AND ANALYSIS SYSTEM

[1]Twinkle Sharma, [2]Pawan Kumar Jaiswal
[1]B.Tech Student, [2]Assistant Professor
[1,2]Department of Computer Science and Engineering
[1,2]Amity School of Engineering and Technology

Abstract

The spread of information is becoming faster due to the advancement of the digital media, as such, users struggle to confirm the authenticity of news article on different platforms. Naradmuni is an intelligent web-based news verification system that focuses on analyzing fake news articles for the benefit of the user. This system takes a hybrid approach; it combines a machine learning method (TF-IDF with Logistic Regression), a rule-based approach and heuristic scoring on the content. A novel feature in Naradmuni is the introduction of "Needs Review" verdict, which reduces overly confident verdict on insufficient information. News articles are fetched in real-time using the News API, which will resort to Google News RSS for a wider range of news and allow users to verify against live reporting. The system generates informative output in terms of the confidence scores, a brief summary and signal that it detects. The user has an efficient system which guides them through the stages of searching, verifying, providing feedback to the system which takes a human-in-the-loop approach and shows that hybrid methods are much safer to use compared to only automated systems for the task..

Keywords: Fake News Detection, Machine Learning, Natural Language Processing, TF-IDF, Logistic Regression, Misinformation, Human-in-the-loop, Explainable AI, News Verification, Hybrid System

1. Introduction

The digital revolution has also changed how we produce, consume, and share information. The increased popularity of social media, instant messaging, and web portals has made information readily available to all, but has also contributed to its spread at an unstoppable rate. The proliferation of online misinformation (referred to as an 'infodemic') poses a real threat to the quality of public discourse, democratic systems, and society as a whole.

Misinformation ranges from completely fabricated news articles, distorted headlines, to out of context but technically true news snippets, doctored visuals, and rumor mill chatter. Although professional fact-checking organizations have emerged and are in place today, fact-checking is a manual process, a slow and reactive endeavor, such that by the time a claim is verified, the same has been transmitted to millions of users already, thereby introducing a speed discrepancy. Naradmuni helps to address this gap by presenting a web based intelligent verification assistant, whose name itself (Narada- a divine sage, known in Hindu mythology

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

for acting as a messenger and communicator) embodies the shift from propagation of unverified information to responsible communication, in addition to offering a 3 tier verification output of 'Likely True', 'Likely False', 'Needs Review', in contrast to a typical binary output from conventional classifiers which fails to capture the essence of 'newsy' information.

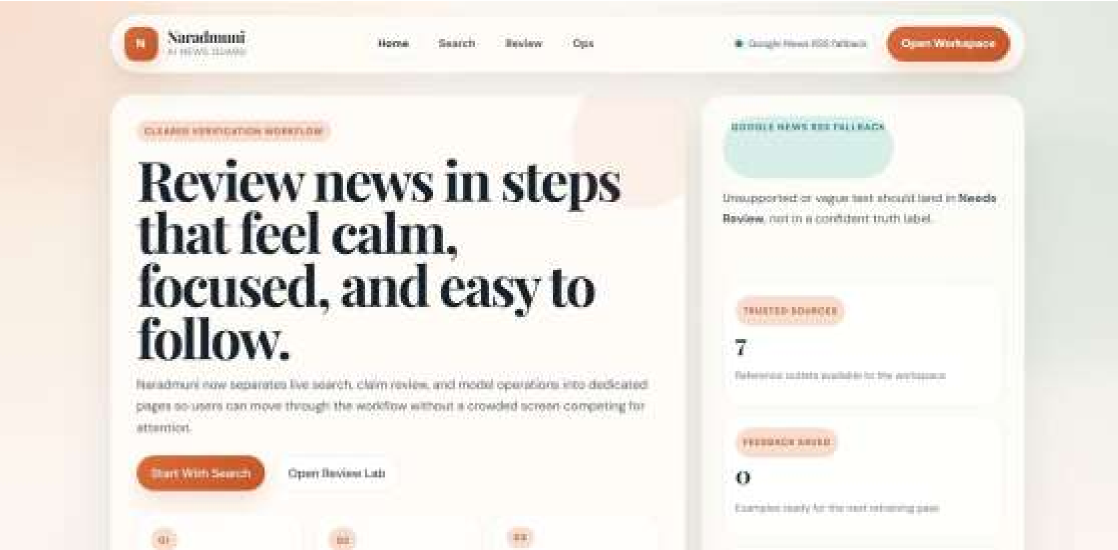

Figure 1: Naradmuni Home Page – Multi-page verification workflow with calm, focused UX design

2. Literature Review

The research into automatic fake news detection has advanced considerably over the last decade on account of the rampant spread of fake news over social media platforms and online news sites. Most of the initial work primarily consisted of supervised machine learning techniques that relied on engineered text features. With progress in NLP and AI, smart and context-aware fake news detection systems are emerging, capable of dealing with voluminous digital content. Today's research goes beyond just detection accuracy and places more importance on explainability, real-time verification and user confidence.

2.1 Traditional Machine Learning Approaches

ML classifiers like Naive Bayes, SVM, Decision trees, Random Forests and Logistic Regression were prevalent in early systems developed for fake news detection. They were feature engineering dependent where relevant patterns from news articles were manually engineered and fed into the classifiers. Some of the widely used feature engineering techniques were: TF-IDF to assign scores to individual words based on their frequency in a document compared to all other documents and N-grams (bigrams were quite effective at recognizing misleading phrases such as 'miracle cure', 'secret truth', 'government hiding' etc.). In addition to that, some other features that were engineered to improve classification were: sentiment polarity, pattern of punctuation marks used in an article, read-ability score and sentence structure. Traditional ML models have some key advantages over deep neural networks in terms of lower computation cost, faster processing, better explainability and higher efficiency at an earlier stage. Nevertheless, limitations with respect to understanding context, sarcasm, relation between two words, etc. And high dependency on hand-crafted features often resulted in a drop in their performance when dealing with continuously evolving news patterns.

2.2 Deep Learning Approaches

Deep learning has proved effective at detecting fake news because automatic feature extraction is possible, and the semantic information of text can be better understood. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs) became popular in detecting long-range dependency of text and local features, respectively. LSTMs are good at handling long-range dependency of text, whereas CNNs can detect local features of deceptive text. Transformer based models such as BERT, RoBERT a and Distil BERT have achieved state-of-the-art performance on fake news detection classification. The attention mechanisms applied in transformer based models allow them to model the dependency between any two words in a sentence so that they perform quite well on the detection of fake news. Transformer based models are particularly capable of detecting subtle misinformation ,ambiguous claims, and context-dependent language. However, deep learning models rely heavily on massive training data, substantial computing power, and robust hardware. Furthermore, deep learning models are difficult to interpret and understand, which hinder people from trusting the prediction from deep learning based models.

2.3 Hybrid Approaches and Research Gap

Recent researches also showed that integrating machine learning models with rule based systems and heuristic scoring methods may enhance both the robustness and interpretability of fake news detection. These hybrid approaches are based on combined methods such as combining natural language context analysis with the scoring mechanisms (credibility, sentiment, and keyword based fact check). Some methods also include fetching news in real time and retrieving additional information to verify the suspicious fact-checked content.

Although many progress have been made, there is still a lack of certain elements of research. Many fake news detection systems still provide just "true" or "fake" as outputs rather than describing the real fact more ambiguously and in context. Real world situations do possess cases of partly true and misleading context dependent news that demands uncertainty based analysis rather than just simple binary classification. Apart from that, a number of present methods do not provide real-time fact verification or clearly explain their verdict. In order to fulfill these gaps, a hybrid framework named as Naradmuni is presented, which uses contextual analysis with heuristic reasoning and real-time fact verification.

3. Methodology

Naradmuni follows a multi-stage hybrid pipeline combining machine learning, heuristic scoring, and rule-based validation. The architecture is designed for efficiency, interpretability, and responsible uncertainty handling.

3.1 Data Collection and Preprocessing

The data set used here is manually crafted, using a CSV file, which includes news with True/False labels. Lowercasing, removing punctuation, stopwords, and tokenization is performed as part of the pre-processing steps. For instance, the input "Government CONFIRMED new scheme!!!" becomes "government confirmed new scheme."

3.2 Feature Extraction: TF-IDF with N-grams

Text is converted to numerical feature vectors using Term Frequency–Inverse Document Frequency (TF-IDF). The formula is defined as:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log(N / \text{df}(t))$$

where TF(t, d) is the frequency of term t in document d, N is the total number of documents, and df(t) is the number of documents containing t. Both unigrams and bigrams are extracted to capture phrase-level misinformation signals.

3.3 Classification: Logistic Regression

Logistic Regression is employed for binary classification due to its efficiency, interpretability, and compatibility with TF-IDF features. The classification probability is computed as:

$$P(y=1 \mid x) = 1 / (1 + e^{-(w \cdot x + b)})$$

The model is trained on the curated dataset, serialized as a .pkl file, and loaded at runtime for real-time inference.

3.4 Heuristic Scoring Engine

This ML probability is then tweaked using a keyword-driven heuristic engine. Positive signals, such as "official", "confirmed" or "research", enhance the credibility value while negative signals, such as "hoax", "viral" or "conspiracy" diminish it. This domain-driven tweak ensures that the system remains robust with respect to unusual situations.

3.5 Rule-Based Safety Validation

An additional rule-based module covers patterns too dangerous for ML to potentially make mistake with. Examples of this would include impersonation with first person authority, e.g. "I am the Prime Minister", for which ML would be avoided altogether and a direct "Likely False" given.

3.6 Uncertainty Handling: The "Needs Review" Verdict

Otherwise, if model confidence score is less than 68%, or the margin between probabilities of classes is less than 15%, or the text is too short, it gives "Needs Review" instead of predicting in either way. This way prevents highly confidence wrong prediction and requires human examination..

3.7 Sentiment Analysis

The shallow lexical sentiment analyser assigns one of three values (Positive, Negative or Neutral) representing the text's tone. This feature can identify the exploitation of emotion-an increasingly common characteristic of misinformation-and will offer a fourth, interpretable, dimension to the result.

4. System Design and Architecture

Naradmuni uses a 4-layer architecture. The layers include: Frontend (HTML/CSS/JS), Backend (Flask API), Intelligence (Fake News Analyzer), and Data (file-based CSV/PKL files). The responsibility of each layer is well defined, thus they can be scaled and maintained separately..

4.1 Layered Architecture

Table 1: Naradmuni System Layers

Layer	Technology	Responsibility
Frontend	HTML, CSS, JavaScript	User interaction, multi-page workspace
Backend	Python, Flask	API routing, request handling, news fetch
Intelligence	scikit-learn, TF-IDF, Logistic Regression	Text analysis, verdict generation
Data	CSV, pickle (.pkl)	Dataset, model, and feedback storage

4.2 Multi-Page Workspace Design

To ensure each page's scope remains on the current job, Live Coverage Search, Claim Review Lab, and Feedback & Model Ops have their own distinct spaces within the interface. This setup is intuitive for new users and decreases mental overhead. Both the current context and recent review state are preserved to minimize duplicated work as users transition through workflows.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

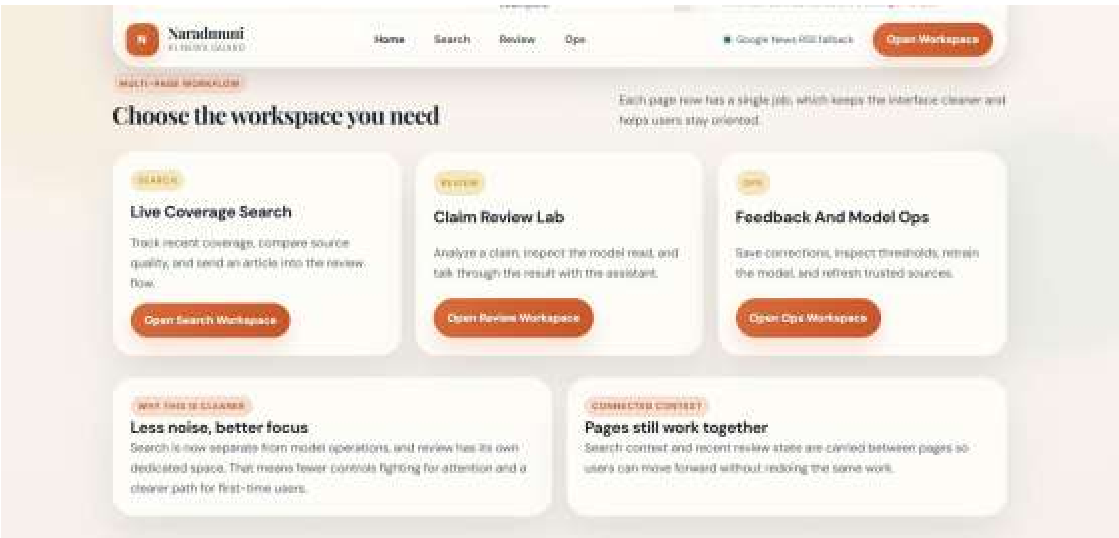

Figure 2: Naradmuni Workspace Selector – Three dedicated workspaces: Search, Review, and Ops

4.3 Live Coverage Search

The Search page allows users to track recent news coverage, compare source quality, and push an article into the review flow. Users can select topic scope (National/Local), set article count, and use pre-defined quick topics such as "Global climate," "Election claims," "Public health," and "AI regulation." A dual-source fallback (News API → Google News RSS) ensures high availability.

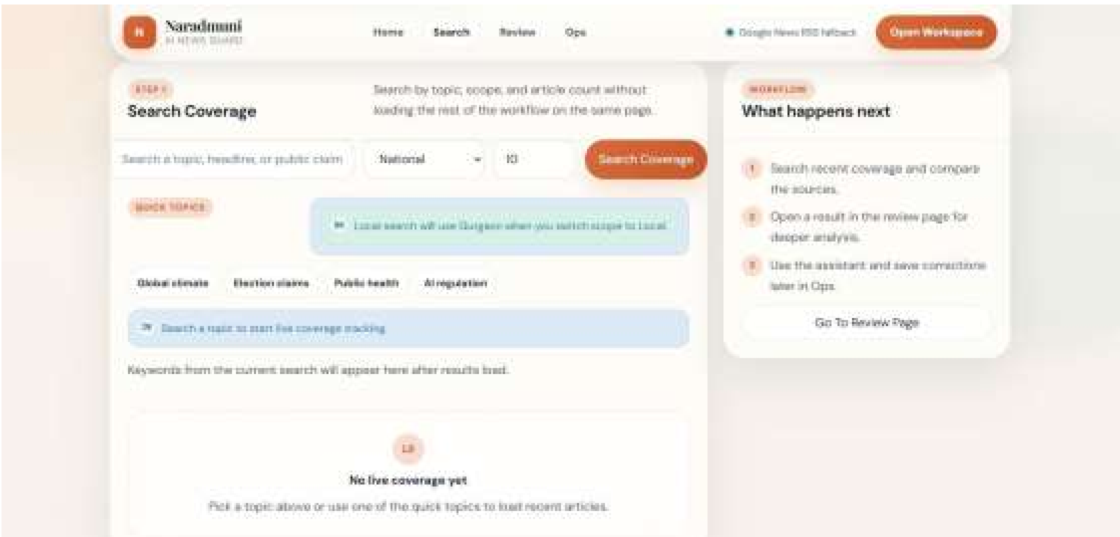

Figure 3: Live Coverage Search Page – Topic-based news search with scope control and workflow guidance

4.4 Key API Endpoints

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

The Flask backend exposes REST endpoints: /news (live news retrieval), /check (claim analysis with verdict and signals), /chat (context-aware conversational guidance), /feedback (user correction storage), /train (on-demand model retraining), and /stats (system monitoring).

5. Results and Discussion

The Naradmuni prototype was evaluated through functional testing, unit testing of the analyzer module, API validation, and edge-case testing. The system was deployed locally and tested across a range of input scenarios.

5.1 Verdict Classification: "Needs Review" in Action

The three-tier verdict system is the centerpiece of Naradmuni's responsible design. When a claim lacks sufficient evidence—such as being too short, lacking source attribution, or having low model confidence—the system returns "Needs Review" with detailed metadata rather than forcing a potentially harmful binary decision.

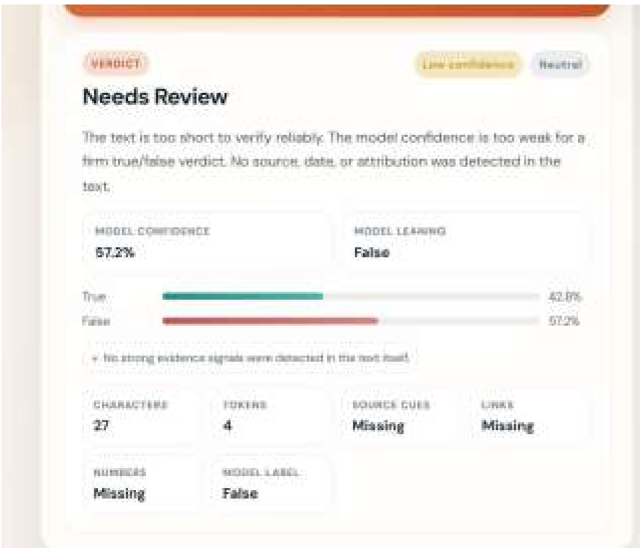

Figure 4: Verdict Panel – "Needs Review" verdict for a low-confidence, short input showing confidence scores, model leaning, and detected signals

5.2 Verdict Classification Performance

Table 2: Sample Verdict Comparison – ML Only vs. Naradmuni Hybrid System

Input Example	ML Only Verdict	Naradmuni Hybrid Verdict
"Miracle cure found for cancer"	Uncertain	Likely False
"Official report released by WHO"	Moderate confidence	Likely True

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

"Check this" (short input)	Low confidence prediction	Needs Review
"I am the Prime Minister..." (impersonation)	Uncertain	Likely False (rule override)

5.3 Discussion Assistant

The integrated Discussion Assistant (Step 2 in the Review workflow) allows users to ask follow-up questions about the current claim or saved search context. Pre-loaded quick questions guide users through interpreting results, understanding the "Needs Review" verdict, and learning how to use trusted sources effectively.

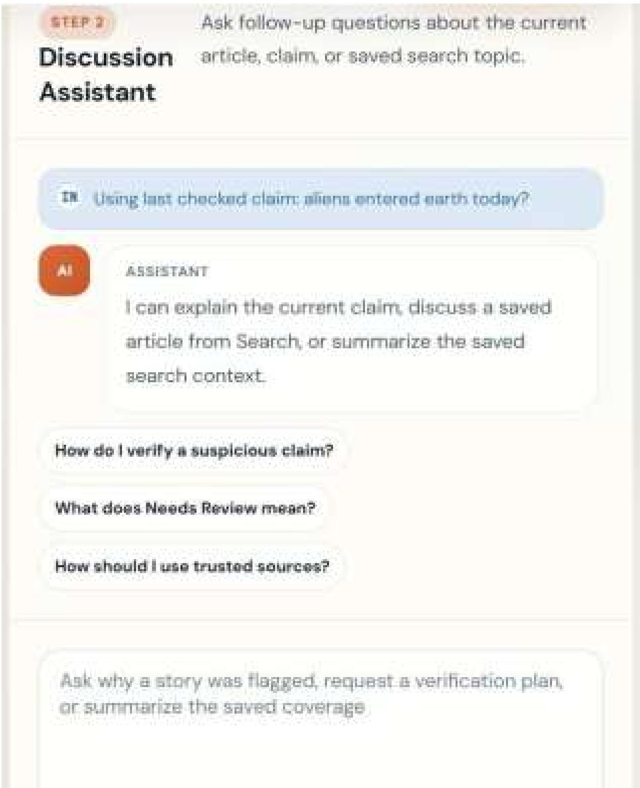

Figure 5: Discussion Assistant – Context-aware chat interface for claim explanation and verification guidance

5.4 System Performance

The system achieved an average API response time of under 1 second on local hardware (Intel Core i3, 4 GB RAM), demonstrating the efficiency of the lightweight TF-IDF + Logistic Regression pipeline. All API endpoints returned correct HTTP status codes and well-formed JSON responses in testing. The "Needs Review" threshold (minimum 68% confidence, 15% probability margin) correctly triggered for short, ambiguous, or incomplete inputs, preventing overconfident misclassification.

5.5 Hybrid vs. Standalone Comparison

Table 3: Naradmuni vs. Traditional Systems

Feature	Traditional Binary Systems	Naradmuni
Output Classes	True / False	True / False / Needs Review
Explainability	Low	High (signals, confidence)
User Guidance	Minimal	Structured workflow + chat
Uncertainty Handling	Absent	"Needs Review" verdict
Real-time Context	None	Live news retrieval
Feedback Loop	None	User correction + retraining

Table 4: Confidence Threshold Decision Table – Logic governing verdict assignment

Condition	Threshold	Resulting Verdict
Model confidence	≥ 68% + margin > 15%	Likely True / Likely False
Model confidence	< 68% OR margin < 15%	Needs Review
Input length	< 10 tokens	Needs Review (too short)
Rule match	Impersonation / high-risk phrase	Likely False (immediate override)

6. Conclusion

Naradmuni successfully proved that the intelligently designed hybrid model is effective even better than individual machine learning models in real-world fake news detection application. TF-IDF + Logistic Regression is combined with heuristic scoring, rule-based safety tests, and a three-tier verdict to produce a usable and responsible detection assistant in the real world. The implementation of the "Needs Review" verdict makes the project more practical as it reflects the inability to definitively label the content rather than committing to an irresponsible binary classification. Adding live news checking, explained outputs, and human-in-the-loop feedback mechanism further separates Naradmuni from previous works. Future works would focus on:

Expanding the training set with multilingual, more diverse sources.

Integrating transformers-based models (BERT/RoBERTa) in our approach.

Developing multilingual support in Hindi and other Indian languages.

Detecting misleading information in images and videos.

Deploying the application on a cloud service to enable scalability in real world.

7. Acknowledgments

We would like to thank Mr. Pawan Kumar Jaiswal (Department of Computer Science & Engineering, Amity University Chhattisgarh) for his immense help and continued support throughout this project. We also like to thank all the open source technologies namely Flask, scikit-learn and Python itself whose developed codes have been put into use in this project.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 1003-1012

References

[1] 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.

[2] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys, 53(5), 1–40.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019.

[4] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In Proceedings of CIKM 2017.

[5] Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. In Proceedings of ACL 2017.

[6] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[7] Oshikawa, R., Qian, J., & Wang, W. Y. (2020). A survey on natural language processing for fake news detection. In Proceedings of LREC 2020.

[8] Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. In Proceedings of ASIST 2015.