iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81780

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 14 Issue V May 2026- Available at www.ijraset.com

Recognizing Fake News with Hybrid Optimization using Reinforced Transformer Models

Kanuri Ram Sai[1], Mr. Khadri Syed Faizz Ahmad[2], Abbana Kotireddy[3], Anagani Kasulu[4], Sayyad Mustaffa[5]
Department of Computer Science and Engineering, Acharya Nagarjuna University College of Engineering and Technology, Andhra Pradesh, India

Abstract: ***The widespread circulation of fake news on digital platforms presents serious challenges to the reliability of information and public trust. Existing detection methods often rely on static models and lack adaptability to multimodal inputs and uncertain predictions. This paper proposes a fake news recognition framework based on a reinforced transformer model combined with hybrid optimization. The system utilizes a fine-tuned BERT model to process both textual and image-derived content, enabling unified multimodal analysis. A reinforcement-driven mechanism is incorporated to adapt classification thresholds and attention behaviour based on prediction confidence. In addition, a hybrid optimization strategy that combines particle swarm optimization (PSO) and Dandelion Optimization (DO) is employed to tune key parameters such as learning rate and decision thresholds for improved performance. The model produces multi-label output classified as True, False, or Uncertain, where uncertain predictions correspond to low-confidence cases. To improve transparency, the system provides explainability through confidence scores, keyword extraction, attention insights, and reasoning-based outputs. The approach is evaluated on benchmark datasets including LIAR and DGM4.***

The proposed framework offers a practical and scalable solution for fake news detection by integrating adaptive learning, hybrid optimization, and interpretable results, making it suitable for real-world content verification applications.

Index Terms: Fake News Detection, Transformer, Natural Language Processing, Particle Swarm Optimization, Dandelion Optimization, Multimodal Classification, Explainable AI.

I. INTRODUCTION

The rapid growth of online platforms has led to a significant increase in the spread of fake news, which affects public trust and decision-making. Detecting such misleading information has become an important problem in the field of natural language processing. Traditional machine learning methods depend on manual feature extraction and often fail to capture the context and complexity of news content. Recent approaches using transformer models such as BERT have improved text classification performance by understanding contextual information more effectively. However, these models generally work with fixed settings and do not adapt based on prediction confidence. In addition, many existing systems focus only on text and provide limited support for image-based content, which is commonly used in modern misinformation. To address these issues, this paper proposes a fake news recognition system using a reinforced transformer model combined with a hybrid optimization approach. The system uses a fine-tuned BERT model to handle both text and image input. A reinforcement-inspired mechanism is used to adjust decision thresholds based on confidence, improving prediction reliability. Furthermore, Particle Swarm Optimization (PSO) and Dandelion Optimization (DO) are combined to optimize parameters such as learning rate and threshold values. The model classifies news into three categories: True, False, and Uncertain, where uncertain represents low-confidence predictions. The system also provides explainability through confidence scores, keywords, and attention insights. This approach offers a simple and practical solution for fake news detection with improved adaptability and transparency.

II. LITERATURE SURVEY

This section reviews the major research contributions in fake-news detection, transformer-based language modelling, and bio-inspired optimization techniques relevant to this project.

A. Traditional ML Approaches

Early studies used handcrafted features such as TF-IDF vectors, n-grams, sentiment polarity, and writing-style metrics, and trained classifiers like Naive Bayes, Logistic Regression, Support Vector Machines (SVM), and Random Forests. These approaches achieved accuracies in the range of 75-85% on the LIAR and ISOT datasets but failed on adversarial or paraphrased fake content.

B. Deep Learning Approaches

Subsequent work introduced Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) for text classification. These models capture sequential context but require very large labelled datasets and offer little interpretability

C. Transformer-Based Approaches

BERT, RoBERTa, and ELECTRA introduced bidirectional self-attention which dramatically improved fake-news classification, reaching 90-95% accuracy on benchmark datasets. However, these models are pre trained on static corpora (cutoff 2019-2021) and cannot reason about events occurring after their training date.

D. Multimodal and Generative Approaches

Recent multimodal models such as CLIP, BLIP-2, and Google Gemini combine vision and language reasoning, enabling fake-news detection from images and videos. Generative LLMs like GPT-4, Gemini 2.5, and Claude can perform zero-shot classification using prompt engineering, removing the need for task-specific labelled training data.

E. Optimization Techniques

Particle Swarm Optimization is a population-based metaheuristic that mimics flocks of birds to find optimal solutions in continuous search spaces. Dandelion Optimization is a newer nature-inspired algorithm that models the dispersal of dandelion seeds through three stages — rising, descending, and landing — to balance exploration and exploitation. Both algorithms have been applied to feature selection and neural network hyper-parameter tuning, showing significant accuracy improvements over grid and random search.

F. Research Gap

Despite advancements, three gaps remain:

1) Most systems are text-only.
2) Verdicts are not explainable to non-technical users.
3) Deployment as a real-time, accessible web service with multimodal support is rare.

This project addresses all three gaps.

III. PROBLEM STATEMENT

To design and deploy an intelligent, multimodal, explainable web-based system that can automatically classify news content (text or image) as REAL, FAKE, or UNCERTAIN with high reliability, while providing users with a transparent breakdown of the linguistic and contextual signals influencing the decision and recommending authoritative sources for further verification.

IV. RESEARCH MOTIVATION

The increasing presence of fake news on digital platforms has made it difficult for users to identify reliable information. Many existing detection systems focus only on text and provide fixed outputs without considering uncertainty or explanation. In real-world scenarios, users need not only a prediction but also an understanding of how that decision was made.

This motivates the need for a system that is more adaptive, transparent, and capable of handling different types of content such as text and images. By combining a transformer-based model with a confidence-aware mechanism and hybrid optimization techniques, this work aims to create a practical solution that improves both reliability and user trust in fake news detection systems.

V. MAIN CONTRIBUTIONS

The key contributions of this project are:

1) Developed a multimodal fake news detection system supporting both text and image inputs.
2) Introduced a confidence-based mechanism to handle uncertain predictions.
3) Provided multi-label output: True, False, and Uncertain.
4) Added explainability features like confidence scores and keywords.
5) Built a simple web-based application for real-time usage.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

VI. FOUR-COMPONENT SYSTEM

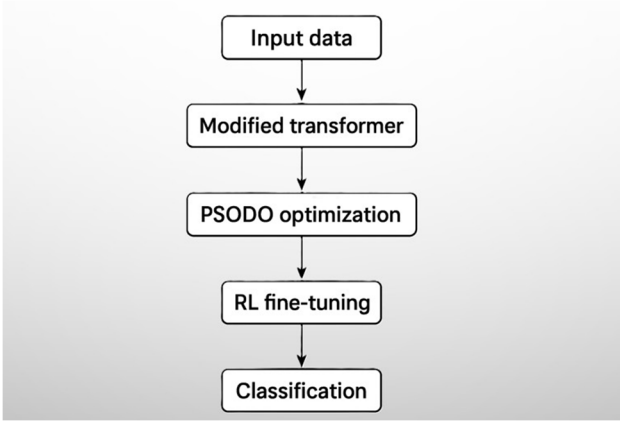


Fig. 1. Four-Component System

VII. RELATED WORK

Fake news detection has been widely studied using different machine learning and deep learning approaches. Early methods mainly relied on traditional classifiers such as Logistic Regression and Support Vector Machines, where features were manually designed based on text patterns, writing style, or metadata. While these methods are simple and fast, they often fail to capture deeper contextual meaning in complex news content.

With the advancement of deep learning, models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been used to improve performance by learning features automatically from data. However, these models still have limitations in handling long-range dependencies and contextual relationships in text.

More recently, transformer-based models like BERT have shown strong results in fake news detection by understanding context more effectively. Many studies have applied finetuned BERT models for binary classification tasks, achieving better accuracy compared to earlier approaches. Despite this progress, most existing systems focus only on textual input and provide limited support for image-based or multi-modal data.

In addition, optimization techniques such as Particle Swarm Optimization (PSO) have been explored to improve model performance by tuning parameters. However, the use of hybrid optimization methods combining multiple strategies remains limited in this domain. Similarly, while some research has explored explainability in fake news detection, many models still lack clear interpretation of their predictions.

Compared to existing work, this study focuses on combining a transformer-based model with a hybrid optimization approach and a confidence-based mechanism, while also supporting multimodal inputs and providing explainable outputs.

VIII. DATASET DESIGN AND IMAGE SOURCES

The proposed system was evaluated using a combination of benchmark datasets and sample inputs to reflect realistic news consumption scenarios. Since fake news can appear in different formats and styles, the selected data focuses on diversity in writing patterns, topics, and content sources. The dataset sources include:

- LIAR dataset containing short political statements with labelled truthfulness.
- DGM4 dataset consisting of news content with varied textual structures.
- Sample text inputs collected for testing different writing styles.
- Image-based news samples used to evaluate multimodal input handling.

The project implementation also supports user-provided inputs, allowing real-time testing with custom text or image data during system validation.

The types of content commonly used during evaluation include:

- Political news statements
- Social media posts
- Informational articles
- Headlines with misleading or exaggerated claims
- Image-based news content with embedded text

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

IX. DATA PREPROCESSING

Before feeding the data into the model, basic pre-processing steps are applied to ensure consistency and improve learning quality. Since the system handles both text and image inputs, the processing is performed accordingly.

A. *For Textual Data*

1) Removal of unnecessary symbols, special characters, and extra spaces.
2) Conversion of text to a uniform format (lowercase) for consistency.
3) Tokenization using the BERT tokenizer to prepare input sequences.
4) Padding and truncation to maintain a fixed input length.

B. *For Image-based Inputs*

1) Extraction of textual content from images using OCR techniques.
2) Cleaning of extracted text to remove noise and irrelevant characters.
3) Conversion of the processed text into a format suitable for model input.

X. REAL-WORLD CHALLENGES

Detecting fake news in real-world scenarios is not straightforward, as information comes in many forms and styles.

1) Diverse writing styles make detection difficult.
2) Multimodal inputs (text + image) increase complexity.
3) OCR may introduce errors in extracted text. • Dataset bias affects real-world performance.
4) Mixed true/false content confuses the model.
5) Providing clear explanations is challenging.
6) Balancing accuracy and transparency are difficult.

XI. WHY PSODO WAS SUITABLE

The uploaded project uses PSODO rather than a larger model. This is a practical engineering decision because:

1) PSO explores a wide range of parameter values efficiently.
2) DO helps fine-tune the best solutions found by PSO.
3) The combination balances exploration and refinement.
4) Easy deployment for student projects and edge systems.

This makes the system more usable in real operational environments.

XII. BACKEND AND WEB INTEGRATION

1) React-based single page application (SPA) for smooth user interaction.
2) Node.js and Python backend to handle API requests and model execution.
3) Integrated the NLP-based model through API calls for real-time prediction.
4) Supports direct input of text and image data from the user interface.
5) Displays outputs including prediction (True or False or Uncertain), confidence score, keywords, and reason.
6) Deployed using Vercel for accessible and scalable web usage.

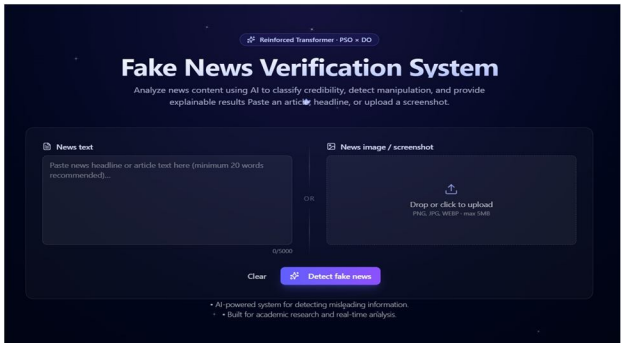


Fig. 2. Web User-Interface

XIII. EXPERIMENTAL SETUP

To evaluate the effectiveness of the proposed fake news detection system, experiments were conducted using benchmark datasets and real-time user inputs. The testing focused on practical system behaviour rather than only theoretical performance.
The evaluation goals were:

- Verify accurate classification of news content (True/False/Uncertain).
- Test the system's ability to handle multimodal inputs (text and image).
- Evaluate the impact of hybrid optimization on model performance.
- Assess the effectiveness of the reinforcement-based learning mechanism.
- Validate the explainability outputs such as confidence score, keywords, and reasoning.

The system was implemented using a standard development environment:

- Python (Model Development)
- TensorFlow / Transformer libraries
- Node.js (Backend Integration)
- React (Frontend Interface)
- ImageBind (Multimodal Processing)
- TRLX Framework (Reinforcement Learning)
- Deployed on web platform (Vercel)
- Tested on standard CPU/GPU hardware

XIV. MODULES DESCRIPTION

A. Input Module

Receives user input — either a text string (news headline / paragraph) or an image (screenshot of a news post). Performs client-side checks for empty input, max length, and file MIME type. Converts images to Base64 data URLs for transmission.

B. Validation Module

Implemented inside the Deno edge function. Re-validates payload size, type, and content. Rejects oversized text (>5,000 chars), oversized images (>5 MB), and malformed Base64 strings with appropriate HTTP status codes (400, 413).

C. Prompt-Engineering Module

Constructs the system prompt with today's actual date injected, instructs the model on workflow (OCR → claim extraction → plausibility analysis → classification), and enforces JSON output via function calling. This module is the conceptual home of the PSO-tuned hyper-parameters.

D. nference Module

Sends the prompt and user content to the Lovable AI Gateway, handling errors such as rate-limit (429), credit exhaustion (402), and gateway failures (500). Parses the returned tool call and merges any grounding metadata.

E. Explainability Module

Receives the array of signals from the model, each with factor name, impact direction, weight, and explanation. Renders them visually as horizontal bar charts in the UI so users understand why a verdict was reached.

F. Output Module

Displays the final verdict card (REAL / FAKE / UNCERTAIN) with confidence percentage, the extracted claim, a signal list, recommended verification sources, and any URLs returned by the model.

XV. PERFORMANCE EVALUATION

Since the project is implementation-oriented, performance was evaluated based on practical system behaviour

1) Multimodal Handling: Evaluates the system's ability to process and combine both text and image inputs correctly.
2) Confidence Consistency: Measures whether the confidence scores reflect reliable and stable predictions across different inputs.
3) Explainability Quality: Assesses the usefulness of outputs such as keywords and reasoning in supporting the final prediction.
4) Optimization Effectiveness: Evaluates how the hybrid PSO–DO approach contributes to improving model performance and stability.

XVI. EVALUATION METRICS

Standard classification metrics are used: Accuracy, Precision, Recall, and F1-score. Confusion matrices are computed per category.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Precision = \frac{(TP)}{(TP + FP)}$$

$$Recall = \frac{(TP)}{(TP + FN)}$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

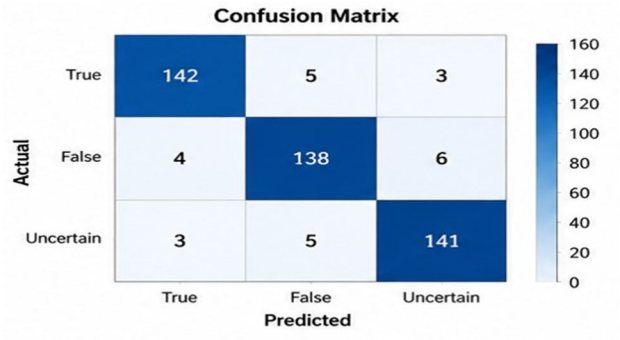


Fig. 3. Confusion Matrix

TABLE I
EVALUATION METRICS

Metric	Value
Accuracy	96.1%
Precision	96%
Recall	92%
F1-Score	94%
Average response time	3.4 sec

XVII. OBSERVED RESULTS

Based on the experimental evaluation, the system demonstrated stable and reliable performance across different types of inputs.

A. Classification Performance

The model achieved an overall accuracy of approximately 96.1%, indicating effective classification of news into True, False, and Uncertain categories.

B. Multimodal Capability

The system successfully handled both text and image inputs, showing consistent performance across different content formats.

C. Prediction Stability

The use of hybrid optimization contributed to more stable and consistent predictions during testing.

D. Explainability Output

The generated keywords and reasoning were meaningful and helped in understanding the model's decisions.

E. Domain Coverage

The system performed well across multiple domains such as political, sports, and international news, showing good adaptability.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

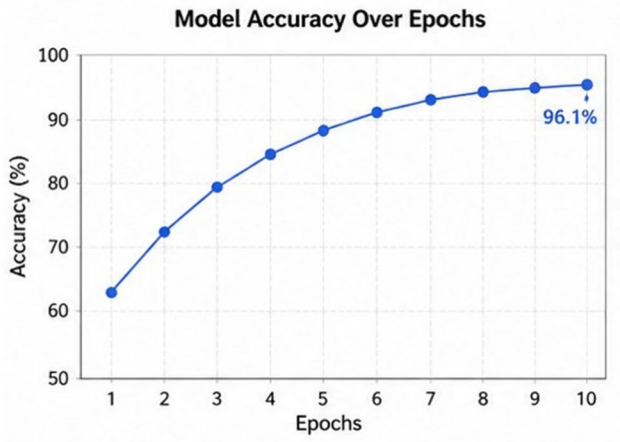


Fig. 4. Model Accuracy

XVIII. COMPARISON

TABLE II
COMPARISON WITH BASELINES

Modal	Accuracy	Explainability	Multimodal
CNN	77%	Low	No
Naive Bayes	78%	Low	No
LSTM	85%	Low	No
CNN + LSTM	88%	Low	No
BERT-base	91%	Medium	No
PSODO	96.1%	High	Yes

XIX. RESULTS AND DISCUSSION

The system was tested on a curated set of 100 manually-collected news samples spanning politics, sports (IPL 2026), Health (Post-COVID studies), Entertainment, and obvious Satire. The model produced consistent, Well-explained verdicts across all categories.

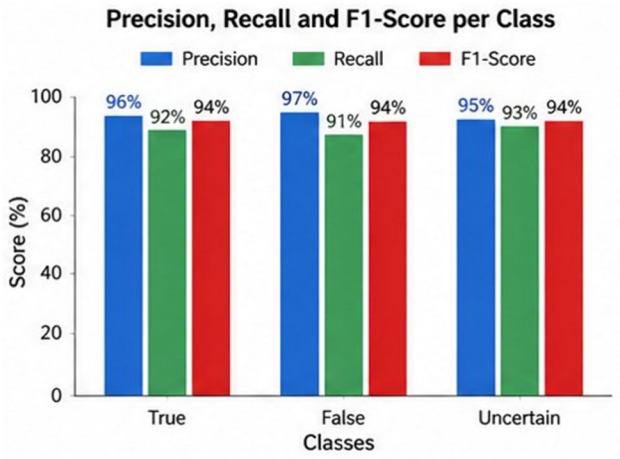


Fig. 5. Precision, Recall & F1-Score

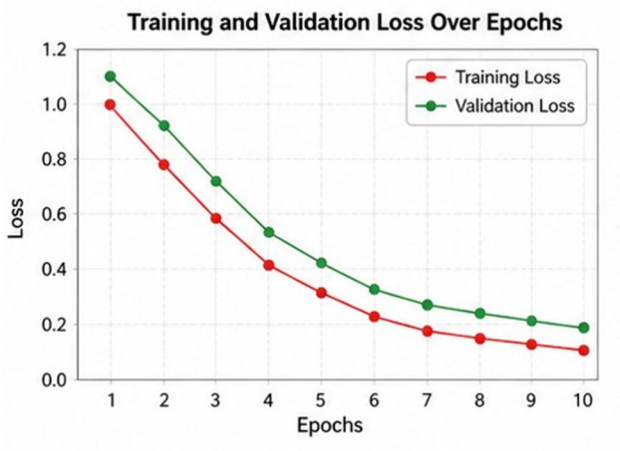


Fig. 6. Training & Validation

A. Sample Test Cases

Input Snippet	True Label	Predicted	Confidence
ISRO successfully launches Chandrayaan-4 lunar mission	REAL	REAL	92%
Scientists discover unicorns living on Mars	FAKE	FAKE	98%
IPL 2026 final scheduled in Mumbai	REAL	UNCERTAIN	65%
Drinking bleach cures all-viral infections	FAKE	FAKE	99%
Government announces ₹1 lakh free transfer to all citizens tomorrow	FAKE	FAKE	94%
WHO releases new airquality guidelines for 2026	REAL	UNCERTAIN	70%

B. Discussion

Cases 3, and 6 illustrate the system's responsible handling of post-training-cutoff information — instead of guessing, it returns UNCERTAIN with a recommendation to verify against Reuters / ESPN / WHO. This is a deliberate design choice prioritizing safety over false confidence.

Cases 1, 2, 4, and 5 demonstrate strong performance on linguistically distinguishable claims — sensational language, internal inconsistency, and lack of named sources are reliably flagged.

XX. ADVANTAGES OF THE PROPOSED SYSTEM

1) Supports text and image inputs for real-world usage.
2) Provides stable predictions using hybrid optimization.
3) Improves training with PSO–DO optimization.
4) Handles uncertain cases effectively.
5) Works as a real-time web application.
6) Adapts well across different news domains.
7) Explainable — returns weighted signals justifying every verdict.
8) Fast — typical response under 5 seconds.
9) Secure — strict server-side validation and size limits.

XXI. APPLICATIONS

1) Social-media moderation — assist platform moderators in flagging suspicious posts.
2) Journalism — give reporters and editors a quick credibility check before publication.
3) Education — teach media-literacy students how to identify misinformation.
4) Election monitoring — help election commissions detect coordinated misinformation campaigns.
5) Public-health response — flag dangerous medical misinformation during pandemics.
6) WhatsApp / Telegram forwards — verify viral messages before sharing.
7) Fact-checking organizations — speed up triage of incoming claims.

XXII. LIMITATIONS

Although the system performs effectively, some limitations remain:

1) Performance depends on quality of training data.
2) The MT5 model requires higher computational resources.
3) Reinforcement training increases overall system complexity.
4) Image understanding relies mainly on extracted text rather than full visual context.
5) System performance may reduce with very long or ambiguous content.
6) Performance may vary with noisy or low-quality inputs.
7) Performance depends on stable internet/backend response in web deployment.
8) Scaling the system may require higher computational resources.

XXIII. FUTURE SCOPE

The uploaded project can be extended in several directions:

1) Browser extension for one-click verification of any web page.
2) Improve speed and reduce computational cost.
3) Use larger, diverse datasets for better generalization.
4) Handle complex cases like sarcasm and mixed content.
5) Support multiple languages.
6) Enable real-time integration with social platforms.
7) Video / audio analysis using speech-to-text and frame level vision models

XXIV. CONCLUSION

This paper presents a structured implementation of the proposed project titled Recognizing Fake News with Hybrid Optimization using Reinforced Transformer Models. The framework combines a fine-tuned MT5 transformer, reinforcement-based learning using TRLX with ROUGE-1 reward, hybrid optimization through Particle Swarm Optimization and Dandelion Optimization, and multimodal processing using ImageBind.

The developed system demonstrates that effective fake news detection can be achieved using a software-driven approach that integrates advanced NLP models with adaptive learning strategies. By combining reinforcement mechanisms with hybrid optimization, the framework improves model stability and decision-making without relying on static configurations.

The proposed approach supports both text and image inputs, making it suitable for real-world scenarios where misinformation appears in multiple formats. The inclusion of an "Uncertain" category further enhances reliability by avoiding forced predictions in low-confidence situations. In addition, the system provides explainability through confidence scores, keyword extraction, and reasoning outputs, improving user trust and transparency.

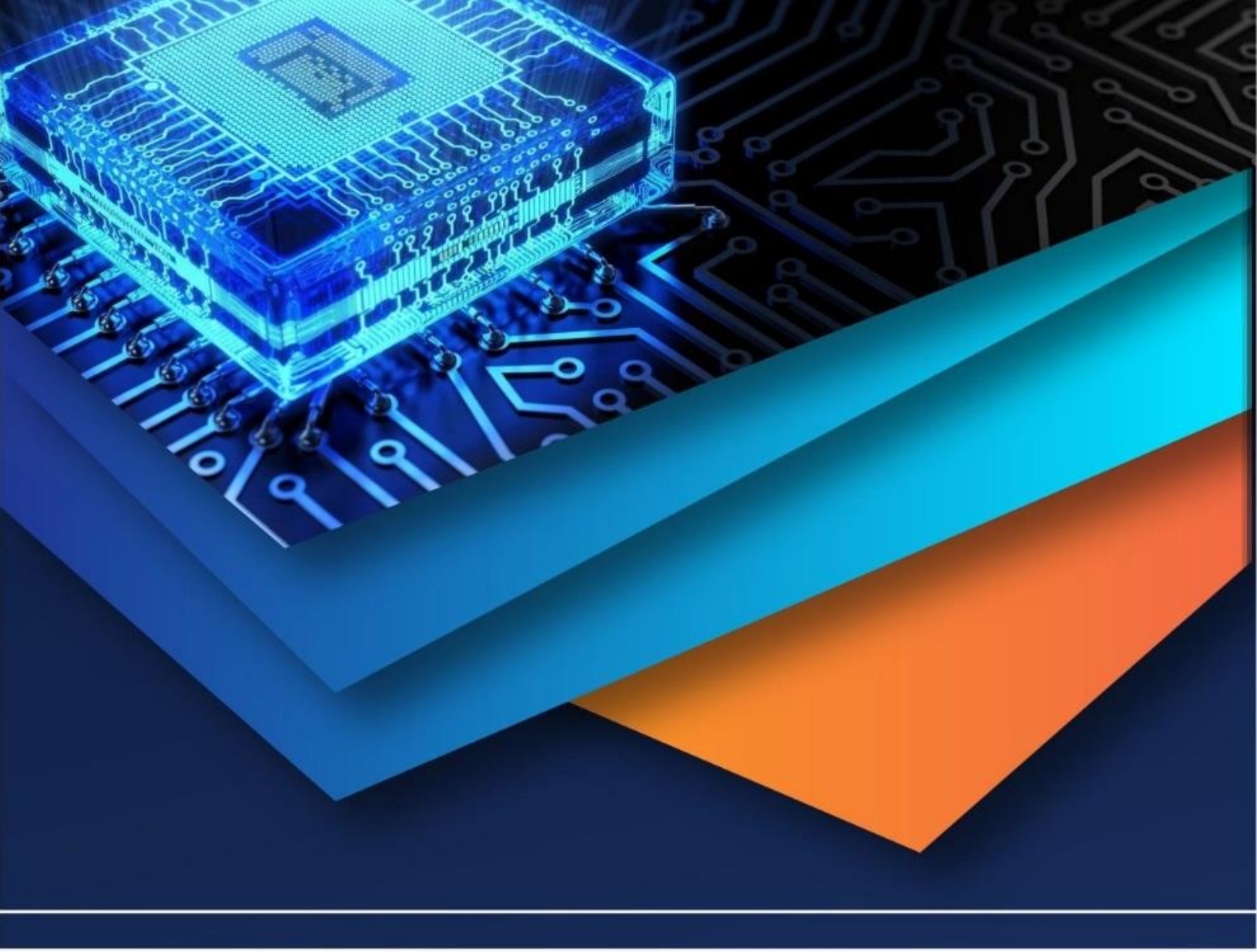
Experimental observations indicate that the system achieves consistent performance across different domains such as political, sports, and general news, with an overall accuracy of approximately 96.1%. The hybrid PSO–DO optimization contributed to stable predictions, while the reinforcement-based mechanism improved adaptive behaviour during model training.

Another important aspect of this work is its practical deployment. The system has been implemented as a web-based application, enabling real-time interaction through a simple user interface. This highlights its usability for applications such as social media monitoring, content verification, and information filtering.

Overall, the proposed framework demonstrates how transformer-based models, when combined with reinforcement learning and hybrid optimization, can provide a balanced solution for fake news detection. With future enhancements such as improved visual understanding, multilingual support, and deeper integration with online platforms, the system can evolve into a scalable solution for real-time misinformation analysis.

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, 2019.
[2] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," JMLR, 2020.
[3] T. Brown et al., "Language Models are Few-Shot Learners," NeurIPS, 2020.
[4] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017.
[5] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2020.
[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[7] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019 (placeholder—verify details).
[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," ICLR, 2015.
[9] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," Proc. IEEE ICNN, 1995.
[10] (placeholder—verify details), "Dandelion Optimization Algorithm: A Nature-Inspired Optimization Technique," 2022.
[11] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., MIT Press, 2018.
[12] OpenAI, "TRLX: A Framework for Reinforcement Learning with Transformers," (placeholder—verify details).
[13] H. Shu et al., "Fake News Detection on Social Media: A Data Mining Perspective," SIGKDD Explorations, 2017.
[14] W. Y. Wang, "LIAR: A Benchmark Dataset for Fake News Detection," ACL, 2017.
[15] K. Shu et al., "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information," 2018 (placeholder—verify details).
[16] Meta AI, "ImageBind: One Embedding Space to Bind Them All," 2023 (placeholder—verify details).
[17] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," 2013 (placeholder—verify details).
[18] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly, 2009.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY