

VeriNews: An Explainable AI Framework for Fake News Detection

Asst. Prof. Manisha Bharatram Bannagare¹

Mr. Kushal Manoj Dhuve², Mr. Junaid Ahmad Khan Javed Ahmad Khan³,

Ms. Kasturi Prashant Deshmukh⁴, Mr. Jay Satishrao Chachane⁵, Mr. Kartik Vilasrao Thakare⁶

Guide, Department of Computer Science & Engineering¹

Students, Final Year Department of Computer Science & Engineering²⁻⁶

R.V. Parankar College of Engineering and Technology, Arvi, Maharashtra

Abstract: *The widespread use of digital news services, social media, and instant communication channels has made information dissemination faster and more accessible, but it has also increased the spread of misleading and fabricated content. This paper introduces VeriNews, an explainable artificial intelligence framework for assisting with the analysis of potentially deceptive news. The framework combines textual processing, contextual language representations, linguistic characteristics, source-oriented signals, external evidence retrieval, and prediction-confidence estimation within a unified verification pipeline. In addition to generating a reliability classification, VeriNews identifies evidence and explanatory factors that allow users to better understand the basis of a model prediction. The proposed workflow includes data collection, text preprocessing, representation learning, classification, evidence retrieval and comparison, confidence assessment, and explanation generation. Benchmark resources such as the LIAR and FakeNewsNet datasets may be employed for model development and evaluation, while transformer architectures including BERT and RoBERTa can be investigated for contextual text representation. The study also considers robustness, dataset-related bias, privacy, interpretability, and the uncertainty inherent in automated information verification. VeriNews is therefore positioned as a human-oriented decision-support framework rather than an autonomous authority on factual truth, particularly in cases where available evidence is limited or inconclusive.*

Keywords: Fake News Detection, Explainable Artificial Intelligence, Natural Language Processing, Fact Verification, Evidence Retrieval, Transformer Models, BERT, RoBERTa, Misinformation Detection, VeriNews.

I. INTRODUCTION

The digital information ecosystem enables news and user-generated content to be published, redistributed, and updated at a very high rate. Information can reach a large audience within minutes, while independent verification may require substantially more time. This imbalance between rapid dissemination and slower verification has created a need for computational approaches that can help identify questionable content and direct users toward relevant supporting evidence.

Detecting fabricated or misleading news remains difficult because deceptive content does not necessarily exhibit obvious linguistic abnormalities. A misleading article may use fluent language, mention real individuals or organizations, cite apparently credible material, or combine factual details with incorrect assertions. As a result, relying only on surface-level keywords or writing-style characteristics may provide insufficient evidence for determining the factual status of a claim.



VeriNews is designed to address this challenge through a multi-stage analysis process that combines content-level signals with contextual information and external evidence. The framework considers three complementary activities: identifying characteristics associated with potentially unreliable content, retrieving information that can be used to examine the underlying claim, and communicating the prediction together with an interpretable confidence measure. By separating these activities, the system is designed to distinguish model-based classification from evidence-based assessment and to make uncertainty explicit when the available information is insufficient for a reliable conclusion.

II. BACKGROUND & RELATED WORK

2.1 Foundations of Fake News Detection

Automated fake news detection has been studied from several perspectives, including the classification of individual claims, complete news articles, and the overall credibility of information sources. Depending on the objective of a system, predictions may be expressed as binary labels, multiple truthfulness categories, or confidence-oriented assessments. Earlier text-based methods primarily relied on lexical characteristics, grammatical structure, sentiment, writing style, and manually designed semantic features. More recent approaches have expanded the analysis by incorporating source credibility, author information, user interactions, propagation patterns, and temporal behavior.

The LIAR dataset is a widely used benchmark for claim-level misinformation research. It contains about 12.8 thousand manually annotated statements gathered from PolitiFact and assigns them to multiple levels of truthfulness together with contextual information [1]. In contrast, FakeNewsNet provides a broader environment for misinformation research by linking news content with social and temporal information. This enables researchers to study not only what is written in a news item, but also how the information behaves and spreads across online platforms

2.2 Traditional and Transformer-Based Methods

Conventional machine learning techniques remain useful for establishing baseline systems in fake news detection. Methods such as TF-IDF combined with Logistic Regression, Naïve Bayes, Support Vector Machines, and Decision Trees can transform textual information into measurable features and provide relatively simple models for comparison. Their computational efficiency and comparatively transparent feature relationships also make them useful for examining whether a dataset contains strong lexical or stylistic patterns.

Transformer-based language models offer a more context-sensitive approach to textual representation. BERT introduced bidirectional contextual encoding through transformer-based pretraining and can be adapted to downstream classification tasks through fine-tuning [3]. RoBERTa further refined the pretraining process by modifying training strategies and optimization choices, leading to stronger contextual language representations [4]. Within fake news detection, such models can be applied to claims, headlines, article bodies, or evidence passages to capture relationships that may not be represented effectively through simple word-frequency features.

2.3 Research Challenges

Despite significant progress, reliable automated misinformation detection continues to face several unresolved issues. The characteristics of misleading information can change as new topics, events, and communication patterns emerge, which may reduce the effectiveness of models trained on older data. Dataset quality is another concern because benchmark collections may contain class imbalance, annotation inconsistencies, duplicated material, or patterns that are strongly associated with particular topics or sources.

Evidence-based verification introduces additional difficulties. Relevant supporting information may be unavailable, difficult to retrieve, outdated, or inconsistent across sources. A model may therefore produce a confident prediction even when the available evidence is insufficient for a definitive conclusion. Another challenge is generalization: a system that performs well on a controlled benchmark may rely on dataset-specific correlations and consequently perform less reliably when evaluated on unfamiliar topics, publishers, or information sources.



Explainability presents a further research challenge. An explanation can appear reasonable to a human reader without accurately reflecting the factors that influenced the underlying model. Therefore, an explainable fake news detection system should not only generate understandable explanations but should also consider whether those explanations are consistent with the evidence and the behavior of the predictive model. These limitations motivate the development of frameworks such as VeriNews that combine classification, evidence retrieval, confidence estimation, and interpretable output rather than relying solely on a final binary prediction.

III. METHODOLOGY

The VeriNews framework is designed to transform an incoming news item into a structured verification report by passing it through a series of interconnected analytical stages. The architecture follows a modular design in which individual components, such as the classification model, evidence retrieval mechanism, or explanation module, can be modified or enhanced without redesigning the complete system.

3.1 System Organization

The proposed architecture is divided into seven major stages: news acquisition, text preprocessing, semantic representation, contextual feature analysis, credibility classification, evidence retrieval and matching, and explanation generation. After processing, the system produces a structured result containing the predicted credibility category, an estimated confidence value, the availability or status of supporting evidence, and a concise explanation of the factors contributing to the prediction.

3.2 News Acquisition and Preprocessing

VeriNews can accept different forms of input, including a news headline, complete article text, individual claim, or web URL. When a URL is provided, relevant textual content can be extracted while excluding unnecessary webpage elements such as navigation components, advertisements, and formatting markup. The preprocessing stage may involve text normalization, sentence boundary detection, tokenization, and removal of duplicated or irrelevant content. Special attention is given to information that can significantly influence the interpretation of a claim. Negation terms, dates, numerical quantities, named entities, and other factual expressions are therefore preserved during preprocessing. Retaining these elements helps prevent the preprocessing stage from removing information that may be important for subsequent verification.

3.3 Feature and Semantic Representation

For a baseline model, the textual input can be transformed into TF-IDF feature vectors and supplied to a conventional machine learning classifier. For a more advanced configuration, a pretrained transformer model can be employed to generate contextual representations in which the meaning of a word is influenced by its surrounding text.

In addition to textual representations, VeriNews may incorporate auxiliary indicators such as similarity between the headline and article body, named-entity characteristics, presence of citations or references, article length, and selected source-level metadata. These variables are considered complementary evidence for the prediction and are not interpreted individually as proof that a news item is either true or false.

3.4 Classification and Confidence Estimation

Following feature extraction, the classification component assigns the analyzed content to predefined credibility categories. Possible categories include likely reliable, potentially misleading, likely false, and requires further review. The exact category structure can be adjusted according to the dataset and experimental requirements.

The system also generates a confidence estimate alongside the predicted category. These values should be obtained from validated model outputs rather than being interpreted directly as certainty. Model calibration is therefore an important part of the methodology, as a probability estimate should reflect the observed reliability of predictions on



suitable validation data. This helps reduce the possibility that a highly confident numerical output is incorrectly interpreted as definitive factual evidence.

IV. VERINEWS ARCHITECTURE AND PROCESSING COMPONENTS

The VeriNews architecture is proposed as a layered verification system in which textual interpretation, claim analysis, evidence search, prediction, and explanation are handled as separate but connected operations. Instead of producing a single label directly from the input article, the framework maintains intermediate information that can be inspected during verification. This organization is intended to make the system easier to evaluate, update, and audit.

The processing pipeline can be represented as:

User Input → Content Extraction → Claim Identification → Signal Construction → Evidence Search → Evidence Assessment → Reliability Estimation → Explanation Report

Each stage produces information required by the following stage, while the final report preserves selected intermediate results for interpretation.

4.1 Content Intake and Normalization Module

The first component receives one of four supported input forms: a news headline, an individual claim, an article text, or a web address. The system assigns an input identifier and converts the received material into a common document structure.

For URL-based input, the extraction process attempts to separate the article body from elements that do not represent the actual report, such as menus, advertisements, repeated navigation text, scripts, and other page-level components. When available, additional metadata such as publisher name, publication date, author, section, and source domain can be stored.

A normalized record may contain the following fields:

Field	Description
document_id	Unique identifier for the analyzed item
headline	Extracted or supplied title
body_text	Main textual content
source_name	Publisher or website name
publication_date	Date associated with the article
author	Author information when available
input_type	Claim, headline, article, or URL

The purpose of this stage is to create a consistent representation before linguistic or verification operations begin.

4.2 Claim Identification Module

A complete news article may contain several statements, but not every sentence needs independent verification. VeriNews therefore introduces an intermediate claim-identification stage that attempts to locate the statements carrying the main factual content of the article.

Candidate claims can be obtained using sentence segmentation followed by a combination of linguistic patterns, named entities, numerical expressions, temporal references, and sentence-level relevance scores. A ranking mechanism can then prioritize statements that contain specific assertions about people, organizations, events, quantities, or relationships.

For an article containing n sentences, each sentence s_i can be assigned a claim priority value:

$$P_i = \alpha E_i + \beta T_i + \gamma N_i + \delta R_i$$



where E_i represents entity density, T_i represents temporal or numerical information, N_i represents linguistic indicators of a factual assertion, and R_i represents relevance to the article topic. The parameters $\alpha, \beta, \gamma, \delta$ can be determined during model development rather than being assumed to be universal constants.

The highest-ranked statements form the initial verification set.

4.3 Linguistic and Contextual Signal Module

Once candidate claims have been identified, VeriNews generates multiple representations of the text. Instead of treating a single feature family as sufficient, the proposed system can combine lexical, semantic, structural, and contextual signals.

The signal groups include:

Lexical signals: term frequency, vocabulary distribution, repeated expressions, and n-gram characteristics.

Structural signals: sentence count, average sentence length, punctuation distribution, and headline-to-body consistency.

Entity signals: names of people, organizations, locations, dates, quantities, and relationships between identified entities.

Semantic signals: contextual vector representations generated using transformer-based encoders such as BERT or RoBERTa. BERT introduced bidirectional contextual pretraining for downstream language tasks, while RoBERTa explored changes to the pretraining procedure for stronger language representations.

Source-context signals: publication date, source domain, available author information, citation presence, and other non-content metadata.

These signals are stored separately so that their contribution can be examined during later analysis. A source-related feature, for example, is not treated as proof that an article is false.

4.4 Hybrid Evidence Search Module

The evidence search component is responsible for finding information that can be used to examine the selected claim. VeriNews can employ a two-step retrieval strategy rather than depending exclusively on keyword matching.

First, a lightweight lexical search can identify documents containing important claim terms, entities, dates, or phrases. The resulting candidates can then be re-ranked using semantic similarity between the claim representation and the retrieved passages.

The proposed retrieval process can be expressed as:

$$R(q, d) = \lambda L(q, d) + (1 - \lambda) S(q, d)$$

where q is the claim, d is a candidate document, L is a lexical relevance score, S is a semantic similarity score, and λ controls the contribution of each retrieval method.

This hybrid strategy is intended to reduce dependence on exact wording. Two statements referring to the same event may use different vocabulary, while semantically unrelated documents can sometimes share many keywords.

Evidence-based fact verification is a well-established research direction. FEVER, for example, contains 185,445 claims and associates supported or refuted claims with evidence sentences, illustrating the importance of separating evidence retrieval from the final verification decision.

4.5 Evidence Assessment Module

Retrieved passages are not automatically considered proof. Each candidate must first be examined for its relationship to the claim.

VeriNews can assign one of three intermediate evidence states:

Supporting evidence: the retrieved information is consistent with the factual content of the claim.

Contradictory evidence: the retrieved information contains information that conflicts with the claim.

Unresolved evidence: the retrieved material is related to the topic but does not provide enough information to establish support or contradiction.



A textual entailment or natural-language-inference model can be used to estimate these relationships. The evidence module can also retain the top-ranked passages so that the final explanation is connected to identifiable source material. This separation is important because fact-checking systems may fail at different points of the pipeline. A wrong conclusion can originate from poor retrieval, incorrect evidence interpretation, or an inappropriate final classification. Automated fact-checking research commonly treats evidence handling as a distinct part of the verification process.

4.6 Reliability Estimation Module

The reliability module combines model prediction and evidence-related information into a structured assessment. Rather than using a single probability as the final truth value, the proposed framework maintains multiple indicators. An illustrative reliability formulation is:

$$V = w_1C + w_2E + w_3Q - w_4U$$

where:

C = classifier confidence,

E = evidence agreement score,

Q = retrieval quality,

U = unresolved or conflicting evidence indicator,

w_1, w_2, w_3, w_4 = parameters learned or selected during evaluation.

The output can then be mapped to operational states such as:

Reliability State	Interpretation
High Support	Model prediction agrees with sufficiently strong evidence
Low Support	Prediction exists but evidence is weak
Conflict Detected	Retrieved sources disagree with the predicted claim
Insufficient Evidence	Available information is inadequate for a reliable conclusion
Review Required	The case should receive human examination

These categories are intended as system states rather than declarations of absolute truth. Automated fact-checking literature has also emphasized the limitations of treating model predictions as complete replacements for human verification.

V. DATASETS AND ANALYTICAL CONSIDERATIONS

The reliability of a fake-news detection study depends strongly on the characteristics of the data used for model development and evaluation. Differences in article length, subject area, source type, annotation procedure, and time period can substantially influence the behavior of a trained classifier. VeriNews therefore treats dataset selection and preparation as an important part of the experimental design rather than as a preliminary step only.

Before model training, the collected records should be inspected for repeated articles, near-duplicate claims, missing values, inconsistent labels, and unusually dominant sources. Training, validation, and test partitions should be constructed carefully so that closely related content does not unintentionally appear in more than one split. Such leakage can make a model appear more accurate than it actually is when it encounters previously unseen information.

5.1 LIAR Dataset

LIAR is a claim-level benchmark created from statements collected from PolitiFact. The dataset contains approximately 12.8 thousand manually labeled short statements and provides multiple truthfulness categories along with contextual information about the statements.

The dataset is particularly suitable for evaluating models that operate on short factual claims. However, its structure is different from that of complete news articles because a short political statement may not contain the same linguistic, structural, or contextual information as a long-form report. Consequently, performance obtained from LIAR should be



interpreted as evidence of claim-classification capability rather than as a direct measurement of performance on every form of online news.

For VeriNews, LIAR can be used to examine the initial claim-analysis and classification components. Its contextual fields can also support experiments on whether auxiliary information improves prediction beyond the textual statement alone.

5.2 FakeNewsNet

FakeNewsNet is a multidimensional repository developed for studying fake news in social-media environments. It combines information associated with news articles with social and spatiotemporal information, allowing researchers to investigate both the content of a news item and aspects of its online dissemination.

The available repository includes samples associated with sources such as PolitiFact and GossipCop, together with fields such as article URLs, titles, tweet identifiers, article text, publication dates, and social-media related information. The maintainers also note that complete social data cannot simply be redistributed because of platform policies and publisher restrictions.

Within VeriNews, FakeNewsNet can support experiments comparing text-only models with models that additionally consider contextual or propagation-related signals. Such information should nevertheless be treated as behavioral context. The presence of high engagement or particular user activity should not itself be interpreted as proof that a news claim is factually correct.

5.3 PHEME for Rumour-Oriented Analysis

A complementary dataset can be obtained from the PHEME family of rumor datasets. PHEME contains social-media conversations associated with breaking-news events and provides annotations related to rumor status and veracity. The dataset is therefore useful for examining cases in which information is incomplete, rapidly changing, or still unresolved. PHEME can be particularly relevant to the uncertainty-handling objective of VeriNews. In addition to distinguishing true and false information, rumor-oriented datasets provide examples where available evidence is not sufficient to establish a definitive conclusion. This property is useful when evaluating the proposed “insufficient evidence” or “review required” states.

PHEME should not be treated as interchangeable with article-level fake-news datasets because its primary setting is social-media rumor analysis. Results obtained from it should therefore be reported separately and interpreted according to the characteristics of the dataset.

5.4 Dataset Comparison for VeriNews

The three datasets provide different evaluation perspectives:

Dataset	Primary Unit	Major Information	Potential Use in VeriNews
LIAR	Short factual statement	Truthfulness labels and contextual attributes	Claim classification
FakeNewsNet	News item and social context	Article, social, and temporal information	News-content and contextual analysis
PHEME	Social-media rumor/conversation	Rumor and veracity information	Uncertainty and rumor verification

Using multiple datasets can provide a broader assessment of the framework because success on one data source does not necessarily imply equivalent performance on a different information environment.



5.5 Data Preparation Strategy

A consistent preprocessing and partitioning procedure should be applied before training. Records with missing critical fields can be flagged for inspection, while duplicate and near-duplicate content should be identified before the dataset is divided into training and evaluation groups.

A recommended experimental arrangement is:

Training Set: used to learn model parameters.

Validation Set: used for model selection, threshold adjustment, and confidence calibration.

Test Set: reserved for final evaluation and not used during model development.

Where the dataset contains timestamps, a temporal evaluation can additionally be considered. In such a setup, earlier records are used for training and later records are reserved for testing. This provides a more realistic assessment of whether a model can handle changes in topics and language over time.

Source-aware splitting can also be useful. When several articles originate from the same publisher or closely related sources, the experiment should assess whether the model is learning generalizable information or simply recognizing source-specific patterns.

5.6 Data Quality and Bias Considerations

Several properties of misinformation datasets can influence the interpretation of experimental results.

Class imbalance: Some credibility categories may contain substantially more examples than others. A model can therefore obtain a misleadingly strong overall accuracy while performing poorly on minority classes.

Annotation uncertainty: Human fact-checking can involve disagreements, particularly when claims depend on context, interpretation, or incomplete information.

Source concentration: If a small number of publishers dominate the dataset, a classifier may learn source-related cues rather than genuine misinformation characteristics.

Topic leakage: Repeated references to the same event, person, organization, or political topic can cause the model to exploit topic-specific patterns.

Temporal drift: The language and subject matter of misinformation can change across time, reducing performance when a model is applied to later or substantially different events.

Duplicate and near-duplicate records: Closely related articles or repeated claims appearing across multiple splits can artificially inflate evaluation scores.

These issues should be documented explicitly because dataset construction can affect the conclusions drawn from the experiment.

VI. EXPERIMENTAL DESIGN

The evaluation of VeriNews is structured to compare conventional text-based classifiers with transformer-based approaches and to examine the effect of contextual and evidence-oriented information. The study focuses on both predictive performance and the usefulness of the resulting system when processing information that was not present during model development.

6.1 Study Objectives

The experimental study is centered on four main objectives. First, it examines the classification performance of the proposed models. Second, it analyzes whether contextual features provide additional information beyond the textual representation. Third, it evaluates the effect of the evidence-matching component on the verification process. Fourth, it investigates whether model behavior remains consistent across different sources and time periods.



6.2 Model Training and Evaluation

A conventional classifier based on TF-IDF features and either Logistic Regression or Support Vector Machine can be used as the baseline configuration. Transformer-based alternatives, including BERT and RoBERTa, can then be fine-tuned using the same data partitions to provide a comparable evaluation.

To study the contribution of individual components, ablation experiments can be performed by removing the contextual-feature component or the evidence-matching stage while keeping the remaining setup unchanged. The resulting differences can then be used to examine the additional contribution of these components.

The available data should be divided into training, validation, and test portions. Model learning and configuration decisions should be performed using the training and validation data, while the test portion should be reserved for the final assessment of the selected models. This separation helps ensure that final performance is measured on data that was not used during model development.

6.3 Performance Metrics

Several evaluation measures are used to describe classification behavior from different perspectives. Accuracy represents the fraction of all predictions that are correct. Precision indicates how many instances predicted as a particular positive category are actually correct, while recall measures the proportion of relevant instances successfully identified.

For multi-class experiments with unequal class frequencies, Macro-F1 can be reported by calculating the F1 score independently for each class and then averaging the resulting values. This prevents a frequently occurring class from dominating the overall summary. A confusion matrix can additionally be used to visualize the distribution of correct predictions and the categories that are most often confused.

Because VeriNews also produces confidence estimates, calibration analysis can be included to examine whether the reported confidence values are consistent with the actual correctness of predictions.

6.4 Robustness and Human Review

The robustness analysis evaluates the models under modified versions of the input. These tests can include paraphrased claims, headline-only content, shortened articles, spelling variations, previously unseen publishers, and test data separated from the training period.

In addition to quantitative evaluation, a sample of generated explanations can be examined by human reviewers. The review focuses on whether the retrieved evidence is genuinely relevant to the claim and whether the explanation provides an appropriate representation of the model's conclusion.

VII. EXPECTED RESULTS AND ANALYTICAL INSIGHTS

The planned experiments are expected to reveal differences between conventional lexical models and transformer-based approaches. Transformer models may provide stronger contextual representations than basic word-based methods, particularly when the meaning of a claim depends on relationships between words and sentences. The addition of evidence retrieval is also expected to provide useful information in cases where the factual status of a claim cannot be determined reliably from linguistic characteristics alone.

The evaluation is also expected to highlight the distinction between prediction confidence and factual correctness. A classification model may assign a high probability to an incorrect prediction, particularly when the characteristics of the evaluation data differ from those observed during training. Consequently, the VeriNews output is designed to contain more than a simple binary classification. The generated report combines the predicted category with its confidence value, the availability of relevant evidence, and an indication of uncertainty.

Another objective is to determine whether incorporating the evidence component helps the system perform more consistently when it is applied to unfamiliar topics or publishers. A substantial reduction in performance on data collected from later time periods would suggest the presence of distributional changes between the development and



evaluation data. Such an observation would highlight the importance of repeated evaluation and model updating rather than relying on a single benchmark measurement.

The expected analysis therefore focuses not only on the numerical classification results but also on how the different components of VeriNews behave when the system encounters information that differs from its training conditions.

VIII. LIMITATIONS AND ETHICAL CONSIDERATIONS

Automated fake-news detection systems cannot establish factual truth solely from linguistic patterns or textual characteristics. Certain claims may depend on domain-specific expertise, original source documents, scientific publications, or up-to-date official information. The evidence retrieval process may also be unsuccessful when appropriate information cannot be found or when reliable sources provide conflicting accounts.

Features derived from the source of a news item may unintentionally create disadvantages for publishers that have less established online presence. Similarly, signals based on social-media engagement can represent the popularity or reach of information rather than its factual accuracy. These factors should therefore be interpreted carefully when they are included in the verification process.

Privacy must also be addressed when user-related or social-media information is collected and processed. The system should avoid gathering personal information that is not necessary for verification. Individual users should not be treated as direct indicators of whether a particular claim is factually correct.

VeriNews is intended to function as an assistive verification and decision-support framework rather than an autonomous authority on factual truth. Human reviewers should have the opportunity to examine the retrieved evidence, challenge system outputs, and identify or correct incorrect predictions. Practical deployment should also involve continued observation of data drift, adversarial attempts to manipulate the system, and changes in the surrounding information environment.

IX. CONCLUSION

This study has introduced VeriNews as an explainable framework for supporting the analysis of potentially misleading news content. The proposed approach brings together natural language processing, contextual information, evidence retrieval, confidence assessment, and interpretable reporting within a unified verification workflow. Its design draws upon established approaches in claim-level classification and social-context analysis while giving particular attention to transparency and the communication of uncertainty.

A key aspect of VeriNews is the separation of model prediction from factual verification. A classification probability is not treated as conclusive evidence about the truth of a claim. Instead, the framework is designed to present the prediction together with the availability and status of retrieved evidence and an understandable explanation, allowing users to examine the basis of the system's assessment.

Future development of the framework can focus on multilingual fact verification, decomposition of complex claims into smaller verifiable statements, retrieval-augmented generation, temporal knowledge graphs, consistency analysis across multiple sources, and more reliable confidence calibration. These extensions can further improve the usefulness of VeriNews for evidence-aware and transparent misinformation analysis.

REFERENCES

1. W. Y. Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vancouver, Canada, 2017, pp. 422–426, doi: 10.18653/v1/P17-2067.
2. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.



3. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
4. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
5. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600.
6. K. Shu, S. Wang, and H. Liu, "Beyond News Contents: The Role of Social Context for Fake News Detection," in *Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM '19)*, Melbourne, VIC, Australia, 2019, pp. 312–320, doi:10.1145/3289600.3290994.
7. J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: A Large-Scale Dataset for Fact Extraction and VERification," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, New Orleans, LA, USA, 2018, pp. 809–819, doi: 10.18653/v1/N18-1074.
8. X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Computing Surveys*, vol. 53, no. 5, Art. no. 109, 2020, doi: 10.1145/3395046.
9. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.
10. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 1321–1330.
11. L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. Wong Sak Hoi, and A. Zubiaga, "SemEval-2017 Task 8: RumourEval: Determining Rumour Veracity and Support for Rumours," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017, pp. 69–76, doi: 10.18653/v1/S17-2006.
12. V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "FANG: Leveraging Social Context for Fake News Detection Using Graph Representation," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*, Virtual Event, 2020, pp. 1165–1174, doi: 10.1145/3340531.3412046.

