A Deep Learning Approach for Disinformation Detection and Forensic Analysis

Vusal Shahbazov[1] Said Mammadov[2]
[1,2]Institute of Information Technology, Baku, Azerbaijan
[1]*v.shahbazov@iit.science.az*, [2]*said.mamadov@gmail.com*

***Abstract*— The spread of disinformation on digital platforms has become a serious threat to social stability, public trust and national security. Timely detection and analysis of such content is currently a pressing challenge in the fields of cybersecurity and digital forensics. This paper presents a deep learning-based approach for automated disinformation detection and forensic analysis. Most existing methods are limited to binary classification, whether content is disinformation or not. The proposed approach goes further by extracting forensic data from content, including name, location and organization. The method was evaluated on both synthetic and publicly available datasets, where it achieved a classification accuracy of 99% on the ISOT dataset and 98% on a corpus generated independently for this study. The results demonstrate that the approach can reliably detect disinformation and provide useful forensic data that can be used by investigators in the decision-making process. This allows the approach to be considered a practical tool for real-world digital forensics applications and contributes to broader research efforts aimed at combating disinformation.**

Keywords*— *disinformation detection, forensic analysis, deep learning, natural language processing, text classification.

I. INTRODUCTİON

It has become significantly more difficult due to the rapid development of digital media to differentiate dependable from fabricated information. Serious threats such as unintentional and deliberate disinformation have converted into serious threats to social stability and public discourse [1]. The facility with which misinformation can spread on social media leads to automated detection not only as an academic pursuit but also as a necessity for being operational [2].

As a result, the consequences are evident. During the COVID-19 pandemic, a parallel "infodemic" was announced by the World Health Organisation, which was a non-regulated surge of health misinformation that caused confusion in public health responses worldwide [3]. Moreover, there was a similar impact on the electoral processes. Fake news has come to form political perceptions and break trust within democratic institutions [4]. Content like that constitutes a direct danger to society, creating urgency for accurate and automatic detection [5]. Meanwhile, large language models (LLMs), over time, further complicate matters by working simultaneously as detection tools and as generators of convincing disinformation at a scale [1].

Concentrating exclusively on detection is not enough. There is an acute need for strategies that go beyond classification to incorporate mitigation, forensic accountability, and prevention [6]. Forensic analysis, tracing narrative origins, identifying manipulated content, and reconstructing propagation pathways, requires systems that operate at the level of individual linguistic and semantic features. Graph-based methods have proven effective for modelling the social structure through which disinformation spreads [7]. Named Entity Recognition (NER) that classifies and identifies entities, for instance, organisations, people, locations, and dates inside unstructured text, is specifically valuable for forensic analysis. Integrating NER into detection pipelines has been shown to significantly improve classification performance [2], and applying post-hoc explainability to named entity decisions, combined with pseudo-anonymizing those entities before training, substantially improves model generalization [8]. NER-based 5W1H identification has further been applied as a structured reliability assessment technique using fine-tuned transformer models [9].

II. RELATED WORK

A Machine Learning

Early automated disinformation detection relied on hand-crafted linguistic and stylistic features fed into classical classifiers. In their paper, Zhou et al. [10] advised a theory-grounded framework which examines news at syntactic, semantic, discourse, and syntactic levels, focusing on forensic and social psychology to show deceptive writing patterns, and more importantly, doing it on the article content alone, without any social propagation data, ensuring early detection is possible before a story has spread anywhere. Qureshi et al. [11] took a different angle, building features from the Twitter interaction networks of a story's spreaders rather than from its content. A 92% and 91% accuracy was achieved by XGBoost on this source-based representation on the Politifact and GossipCop datasets, surpassing Random Forest, SVM, Decision Tree, MLP, and KNN. The combination of these two studies makes a crucial observation, which is that content, feature type, network origin, or style, structures detection performance as much as the choice of algorithm.

B Deep Learning

Hand-crafted features have replaced deep learning with learned representations that encode sequence and context straight from raw text. Kaliyar et al. [12] proposed FNDNet, a deep CNN that learns discriminatory features through stacked layers without domain-specific engineering, achieving 98.36% accuracy on benchmark data. Recurrent and convolutional layers were combined by Nasir et al. [13] into a hybrid CNN-RNN

DOI: 10.25045/forensic.gov.2026.21

model in which the CNNs are for local n-gram patterns, and RNNs are for longer sequential context, and clarified it on the ISO and FA-KES datasets, where it considerably outperformed non-hybrid baselines. However, Das et al. [14] used a different approach, utilising an ensemble of pre-trained language models combined with an analytical algorithm that combines metadata signals, username handles, source domains, and author identities alongside the text. The system reached F1-scores of 0.9892 on a COVID-19 dataset and 0.9156 on FakeNewsNet, and uniquely produced calibrated uncertainty scores for each prediction, a quality useful in forensic settings where knowing how confident a classifier is matters as much as knowing what it predicts.

C. Transformer-Based and Knowledge-Enhanced Approaches

Transformer architectures set a new performance ceiling for detection tasks. Classical CNN/RNN, ML and transformer configurations have been compared by Farhangian et al. [15] across multiple datasets and discovered that context-dependent transformers continuously performed best. An instructive secondary finding is the usage of a transformer as a feature extractor paired with a classical classifier, which inferred better and costs less computationally than end-to-end fine-tuning. Roumeliotis et al. [4] put numbers on the gap, fine-tuned GPT-4 Omni reached 98.6% accuracy against 58.6% for CNNs, while also showing that the smaller GPT-4o mini model performed comparably, which matters for real-world deployment. These gains were grounded by Dar et al. [3] in structured knowledge, blending NER-extracted biomedical entities with a knowledge graph immersed for COVID-19 fake news and authorising the model to cross-reference claims against verifiable facts instead of distributional signals alone. On the risk side, Crothers et al. [16] catalogued the threat landscape posed by neural text generation and argued that trustworthiness, defined in terms of fairness, robustness, and accountability, must be a design requirement from the outset. Chen et al. [5] addressed the deployment challenge by distilling LLM detection capabilities into smaller models, preserving most of the performance at substantially lower inference cost.

D. Named Entity Recognition: Methods and Applications

NER identifies and classifies entity spans, persons, organizations, locations, dates within unstructured text. In disinformation these are precisely the elements most manipulated: invented attributions to real individuals, fabricated organizational affiliations, and misleading geographic claims. Therefore, NER grounds automated detection in the factual architecture of a text rather than its stylistic surface.

BiLSTM-CRF architectures have become the workhorse of deep learning NER: recurrent layers build contextual token representations, while the CRF decoder enforces globally consistent label sequences. Gajendran et al. [17] strengthened the architecture by feeding it both character-level and word-level embeddings, thereby enabling the model to capture morphological and orthographic cues alongside distributional context, and achieved F1-Scores of 89.98 and 90.84 on biomedical benchmarks. A distinct structural challenge relevant to disinformation is nested NER, where one entity name contains another, a person's name inside an organization name, for instance. Shibuya and Hovy [18] handled this by training a CRF to treat inner nested entities as the second-best decoding path within their parent span, then extracting entities iteratively from outermost to innermost. The approach achieved F1-scores of 85.82%, 84.34%, and 77.36% on the ACE-2004, ACE-2005, and GENIA datasets respectively, without adding hyperparameters.

BERT substantially improved NER performance by providing deep contextual representations in place of static embeddings. Ghaddar et al. [19] identified a persistent weakness in BERT-based NER, which they called Name Regularity Bias. Models were classifying entity types based on surface name form rather than context. Context-aware adversarial training, which injects learnable perturbations into entity mentions during training, forced the model to depend on contextual signals and produced significant gains on their diagnostic benchmark. The relevance to disinformation is direct: fabricated content frequently places plausible-sounding names in contextually implausible roles, exactly the pattern a surface-biased model would miss. Berragan et al. [20] showed a practical version of the same lesson for geographic entities: domain-specific transformer fine-tuning achieved an F1 of 0.939, compared with 0.730 for general prebuilt models, a gap that matters given how central location manipulation is to geopolitical disinformation.

Within disinformation research specifically, NER has been applied in several ways. Dahou et al. [2] found that fusing NER features from an AraBERT multi-task architecture improved fake news classification in 5 of 7 Arabic datasets, with an average gain of 1.62%. Dar et al. [3] used a biomedical NER model as the entity-extraction layer in a knowledge-graph pipeline, linking COVID-19 news claims to structured factual evidence. Pan et al. [9] applied BERT-based NER to extract 5W1H elements from news articles as a proxy for credibility, achieving the best score of 65.820% on the reliability subtask at FLARES@IberLEF 2024. González-Silot et al. [8] used SHAP to measure the influence of named entities on classifier decisions and found that pseudo-anonymizing those entities before training improved generalization on external test data by over 44%, evidence that without this step, models latch onto entity surface forms as spurious shortcuts. Taken together, these results position NER as a core analytical layer in forensically grounded disinformation systems, not just a preprocessing step.

III. METHODOLOGY

A. Dataset

This study uses two independent datasets. The first is the widely used ISOT Fake News Dataset, a large collection of real and fabricated news articles. The second is a synthetically generated corpus created specifically for this study. Both datasets have an identical schema (headline, text, label), and the additionally generated dataset contains entity information. The model is trained and evaluated independently on each dataset. Table I presents the statistics of both datasets.

TABLE I. DATASETS STATISTICS

Dataset	Total	Real	Fake	Train	Test
ISOT	44,898	21,417	23,481	35,918	8,980
Synthetic	20,000	9,220	10,780	16,000	4,000

B. *Model Architecture*

The proposed architecture is a two-stage pipeline: a frozen BERT-based NER encoder that produces contextual token-level representations, followed by a trainable BiGRU head that aggregates those representations and outputs a binary classification. Figure 1 provides a schematic overview.

C. *NER-BERT Encoder*

The rationale for using a checkpoint pretrained for named entity recognition is that named entities, such as people, organizations, locations, and geopolitical references, carry a strong discriminatory signal for detecting misinformation. Fabricated articles often misrepresent sources of quotes, invent data, or cite nonexistent entities, so a NER-aware encoder is expected to produce representations more sensitive to such inconsistencies than a standard BERT model. Let d denote the input document. Each document is tokenized, truncated, or padded to a fixed maximum length of L = 512 tokens and fed to the encoder. The output is the last hidden state matrix:

$$\boldsymbol{H = BERTNER(D) \epsilon R\text{^}(L \times H)} \quad (1)$$

This matrix retains position-aware contextual information for every token and is passed directly to the classification head as a tensor of shape (batch, 512, 768).

D. *BiGRU Classification*

The classification is implemented in Keras/TensorFlow. It receives a token sequence matrix as input and processes it through three layers, as summarised in Table II.

TABLE II. CLASSIFICATION HEAD LAYER SUMMARY

Layer	Configuration	Output Shape
Input	BERT last_hidden_state	(batch, 512, 768)
Bidirectional GRU	units=64	(batch, 128)
Dropout	rate = 0.2	(batch, 128)
Dense (Softmax)	units = 2	(batch, 2)

Bidirectional GRU. The BiGRU reads the 512-token sequence in both forward and backward directions. At each timestep t it computes two gates based on the current token x_t and the previous memory $\boldsymbol{h_{t-1}}$:

$$z_t=\sigma(W_z\cdot[h_{t-1}, x_t]+b_z) \quad (2)$$

$$r_t=\sigma(W_z\cdot[h_{t-1}, x_t]+b_r) \quad (3)$$

where z_t controls how much the memory updates and r_t controls how much old memory is discarded. The new hidden state is computed as:

$$\boldsymbol{\tilde{h}_t = tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h)} \quad (4)$$

$$\boldsymbol{h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t} \quad (5)$$

The final hidden states of both directions are concatenated into a summary vector:

$$\boldsymbol{s = [h\_512\text{^}(\rightarrow); h\_1\text{^}(\leftarrow)] \in \mathbb{R}\text{^}128} \quad (6)$$

Dropout. A Dropout layer (p = 0.2) randomly sets 20% of values in s to zero during training to prevent overfitting.

Dense, Softmax. The final layer maps s to class probabilities:

$$\boldsymbol{P(y \mid d) = Softmax(W_o \cdot Dropout(s) + b_o)} \quad (7)$$

$$\hat{y} = \boldsymbol{argmax\, P(y \mid d)} \quad (8)$$

The predicted label ŷ is the class with the higher probability.

Training hyperparameter configuration

All trainable parameters are BiGRU. Before training, the NER-BERT encoder is run once for all training and test splits to extract and store embeddings as in-memory NumPy arrays. The full hyperparameter configuration is presented in Table III.

TABLE III. HYPERPARAMETER CONFIGURATION

Hyperparameter	Value
Optimiser	Adam
Learning rate	1×10^{-3}
Epochs	10
Training batch size	64
Loss function	Sparse Categorical CE
Train / Test split	80% / 20%

The Adam optimizer with default parameters is used. Training lasts 10 epochs with mini-batches of 64 samples, shuffled in each epoch.

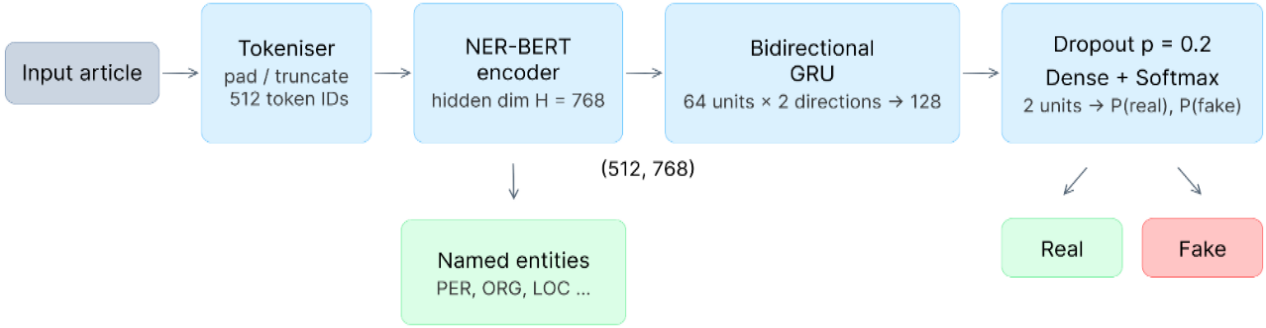

Figure 1. Architecture of proposed pipeline.

IV. RESULTS

A *Classification Performance*

This section presents the classification results of the NER-BERT + BiGRU model evaluated on both datasets. All results are presented for the test set (20%). Table IV summarizes the performance metrics for all experiments using precision, macro-average precision, recall, and F1-score.

TABLE IV. CLASSIFICATION RESULTS ON TEST SETS

Model	Accuracy	Macro Precision	Macro Recall	Macro F1	Dataset
NER-BERT+BiGRU	0.9100	0.9155	0.9193	0.9099	Synthetic
NER-BERT+BiGRU (ft)	0.9837	0.9836	0.9843	0.9837	Synthetic
BERT+BiGRU (ft)	0.9789	0.9782	0.9794	0.9786	ISOT Dataset
NER-BERT+BiGRU (ft)	0.9989	0.9989	0.9988	0.9989	ISOT Dataset

After optimization, performance significantly improves across all metrics, reaching 98.37% accuracy and a Macro F1 score of 0.9837. Both classes demonstrate near-perfect precision and recall. A 7% improvement in accuracy and Macro F1 score was achieved on the synthetic corpus.

The model demonstrates near-perfect results on the ISOT dataset, achieving an accuracy of 99.89% and a Macro F1-score of 0.9989. For both the real and fake classes, precision and recall exceed 0.997, demonstrating the high discriminatory power of NER-BERT representations. This strong performance on ISOT indicates that the named entity information embedded in BERT representations provides a robust signal for distinguishing between real and fake news.

In the ISOT benchmark, replacing the named-entity-aware encoder with the general-purpose BERT base encoder reduces the accuracy from 99.89% to 97.89%, and the Macro F1 score from 0.9989 to 0.9786. This 2% difference, measured on 8,980 test samples, provides direct evidence that named-entity-aware pretraining is superior to standard contextual representations alone. A corresponding comparison on a synthetic corpus has not been completed and remains an open question for future research.

B *Named entity recognition results*

Analysis of the synthetic dataset shows that named entities are recognised at the entity level in 87% of cases, while exact document-level match reaches 62%, confirming that the NER-BERT backbone successfully captures entity-level information in the generated corpus. Table V illustrates an example of entity recognition on a synthetic article for example sentence.

Input sentence: "*Amazon founder Jeff Bezos announced that Blue Origin would launch its next mission from Cape Canaveral in partnership with NASA."*

In this example, 6 out of 6 entities are identified correctly. This is a complete document-level match. The model successfully distinguishes between people (PER), locations (LOC), and organizations (ORG). The baseline NER-BERT model maintains its ability to recognize entities across various domains.

TABLE V. EXAMPLE OF NAMED ENTITY RECOGNITION ON A SYNTHETIC ARTICLE

Predicted Entity	Word
Organization (ORG)	Amazon
Person (PER)	Jeff
PER	Bezos
ORG	Blue Origin
Location (LOC)	Cape Canaveral
ORG	NASA

C *Comparison with state-of-the-art methods*

Table VI presents a comparison of the proposed model with four state-of-the-art detection approaches, evaluated exclusively on the ISOT dataset. All baseline models were selected from recent peer-reviewed publications that explicitly report results on the ISOT dataset under comparable experimental conditions.

TABLE VI. COMPARISON OF FAKE NEWS DETECTION MODELS WITH EXISTING METHODS

Reference	Model	Accuracy	Macro F1	Dataset
[21]	n-gram + Deep Ensemble + MLP	1.000	1.000	ISOT
[22]	Gradient Boosting	0.9961	0.9960	ISOT
[23]	Random Forest	0.9900	0.9900	ISOT
[24]	LSTM + CGPNN + MFWO	0.9200	0.9000	ISOT
Proposed	NER-BERT + BiGRU	0.9989	0.9989	ISOT

The proposed model achieves 99.89% accuracy and a Macro F1 score of 0.9989, ranking second overall and outperforming several baseline approaches. In contrast, the proposed model is the only approach in this comparison refined to recognize named entities. This choice is particularly important for fake news detection, as fabricated articles often misattributed quotes, invent sources, or mention non-existent organizations and individuals. By using a named-entity-aware encoder, the model creates token-level representations sensitive to these entity-level inconsistencies, providing a qualitatively different and complementary signal that purely statistical methods cannot capture.

V. DİSCUSSİON

The NER-BERT + BiGRU model achieves 99.89% accuracy and a Macro F1 of 0.9989 on ISOT, and 98.37% on the synthetic corpus. Named-entity-aware representations provide a strong signal for distinguishing real from fake content. This represents a significant advantage in the context of forensic examination, where detailed information about the object influences investigative decisions.

The ISOT numbers require honest qualification. The dataset carries a well-documented structural artifact: real articles come exclusively from Reuters; fake articles come from a narrow set of politically homogeneous outlets. As a result, a classifier can reach very high accuracy by recognizing source-associated style, not by understanding what makes content false. Every baseline in Table 6 faces the same problem, which explains why several of them also approach or reach 100%. ISOT accuracy is a necessary condition for comparability with prior work. It is not, on its own, evidence of real-world robustness.

The synthetic corpus tells a more conservative story. Because it was generated without source-level artifacts, it strips away the stylistic shortcuts that inflate ISOT performance. The tuned model achieves 98.37%. The gap between the two benchmarks is itself informative: it gives a rough empirical measure of how much ISOT scores are shaped by corpus structure rather than model capability.

The main practical limitation is computational. The NER-BERT encoder contains approximately 110 million parameters and operates on sequences of up to 512 tokens, which requires GPU infrastructure at inference time. For field forensic environments without dedicated hardware, this is a genuine barrier to deployment rather than a theoretical one.

CONCLUSION

This paper presents a deep learning approach for disinformation detection and forensic analysis built on NER-BERT combined with a bidirectional GRU classification head. The contribution has two parts. First, the use of a named entity recognition architecture produces token-level representations that are directly relevant to how real disinformation works: fabricated content routinely misattributes quotes, invents sources, and references individuals or organizations that do not exist. Second, the NER output functions as a forensic layer alongside binary classification, giving investigators structured entity-level evidence rather than a bare label.

The model was evaluated on two independent datasets. On the ISOT benchmark it achieved 99.89% accuracy and a Macro F1-score of 0.9989, outperforming three of four baseline models. These scores should be read alongside the dataset's known source-level bias, which affects every method evaluated on this corpus, as discussed in Section 5. The synthetic corpus, which does not carry those artifacts, gives a more conservative picture: 98.37% accuracy, NER coverage at 87% at the entity level, and 62% exact match at the document level. Across both datasets, the results suggest that NER-aware representations provide a consistent signal for automated disinformation analysis, one that holds beyond the favorable conditions of a single benchmark.

The current model outputs a binary label and a list of detected named entities. A natural and important next step is to generate a structured forensic output. This would take the form of a machine-readable report that includes detected entities, their types, credibility scores, and contextual roles in the document. Such output could be directly used by investigators and digital forensic tools without additional post-processing.

More broadly, the paper reveals a gap between detection and understanding. The model can reliably detect misinformation, but it does not explain what exactly makes content false, what statement is fabricated, what entity is misattributed, or what fact can be verified. Future research should explore what forensic analysts actually need in practice through a structured requirements collection and design the model's output accordingly.

Finally, this approach has significant resource limitations: NER-BERT requires a GPU infrastructure for efficient inference, limiting its use in field forensic environments. Exploring alternative lightweight encoders and optimization methods to bring the system closer to real-world investigative workflows remains an open and practically important task

REFERENCES

[1] S. B. Shah et al., "Navigating the Web of Disinformation and Misinformation: Large Language Models as Double-Edged Swords," IEEE Access, vol. 12, pp. 82729–82756, 2025.

[2] A. Dahou et al., "Linguistic Feature Fusion for Arabic Fake News Detection and Named Entity Recognition Using Reinforcement Learning and Swarm Optimization," Neurocomputing, vol. 596, p. 128078, 2024.

[3] R. A. Dar, R. Hashmy, M. S. Anwar, P. Böhm, and J. Frnda, "Semantic Knowledge Graph Fusion for Fake News Detection: Unifying Content-Based Features and Evidence-Based Analysis in the COVID-19 Infodemic," PLOS ONE, vol. 20, no. 4, p. e0321919, 2025.

[4] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models," Future Internet, vol. 17, no. 1, p. 28, 2025.

[5] X. Chen, X. Huang, Q. Gao, L. Huang, and G. Liu, "Enhancing Text-Centric Fake News Detection via External Knowledge Distillation from LLMs," Neural Networks, vol. 184, p. 107377, 2025.

[6] D. Sallami and E. Aïmeur, "Exploring Beyond Detection: A Review on Fake News Prevention and Mitigation Techniques," J. Comput. Soc. Sci., vol. 8, no. 1, p. 26, 2025.

[7] X. Su, J. Yang, J. Wu, and Z. Qiu, "Hy-DeFake: Hypergraph Neural Networks for Detecting Fake News in Online Social Networks," Neural Networks, vol. 183, p. 107302, 2025.

[8] S. González-Silot, A. Montoro-Montarroso, E. Martínez-Cámara, and J. Gómez-Romero, "Enhancing Disinformation Detection with Explainable AI and Named Entity Replacement," Knowledge-Based Systems, vol. 315, p. 115933, 2026.

[9] R. Pan, J. A. García-Díaz, F. García-Sánchez, and R. Valencia-García, "UMUTeam at FLARES@IberLEF 2024: Enhancing Disinformation Detection with 5W1H Techniques and Transformer Models," Procesamiento del Lenguaje Natural, vol. 73, pp. 333–342, 2024.

[10] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake News Early Detection: A Theory-Driven Model," Digital Threats: Research and Practice, vol. 1, no. 2, pp. 1–25, 2020.

[11] K. A. Qureshi, R. A. S. Malick, M. Sabih, and H. Cherifi, "Deception Detection on Social Media: A Source-Based Perspective," Knowledge-Based Systems, vol. 248, p. 108888, 2022.

[12] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A Deep Convolutional Neural Network for Fake News Detection," Cognitive Systems Research, vol. 61, pp. 32–44, 2020.

[13] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach," Int. J. Inf. Manag. Data Insights, vol. 1, no. 1, p. 100007, 2021.

[14] S. D. Das, A. Basak, and S. Dutta, "A Heuristic-Driven Uncertainty Based Ensemble Framework for Fake News Detection in Tweets and News Articles," Neurocomputing, vol. 491, pp. 607–620, 2022.

[15] F. Farhangian, R. M. O. Cruz, and G. D. C. Cavalcanti, "Fake News Detection: Taxonomy and Comparative Study," Information Fusion, vol. 103, p. 102140, 2024.

[16] E. N. Crothers, N. Japkowicz, and H. L. Viktor, "Machine-Generated Text: A Comprehensive Survey of Threat Models and Detection Methods," IEEE Access, vol. 11, pp. 70977–71002, 2023.

[17] S. Gajendran, M. D., and V. Sugumaran, "Character Level and Word Level Embedding with Bidirectional LSTM – Dynamic Recurrent Neural Network for Biomedical Named Entity Recognition from Literature," J. Biomed. Inform., vol. 112, p. 103609, 2020.

[18] T. Shibuya and E. Hovy, "Nested Named Entity Recognition via Second-Best Sequence Learning and Decoding," Trans. Assoc. Comput. Linguist., vol. 8, pp. 605–620, 2020.

[19] A. Ghaddar, P. Langlais, A. Rashid, and M. Rezagholizadeh, "Context-Aware Adversarial Training for Name Regularity Bias in Named Entity Recognition," Trans. Assoc. Comput. Linguist., vol. 9, pp. 586–604, 2021.

[20] C. Berragan, A. Singleton, A. Calafiore, and J. Morley, "Transformer Based Named Entity Recognition for Place Name Extraction from Unstructured Text," Int. J. Geogr. Inf. Sci., vol. 37, no. 2, pp. 417–435, 2023.

[21] A. Ali, F. A. Ghaleb, B. A. S. Al‐rimy, F. Alsolami, and A. I. Khan, " Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique," Sensors , vol. 22, no. 18, pp. 6970‐6970, Sept. 2022, doi: 10.3390/s22186970.

[22] B. E. Elbaghazaoui, M. Amnai, Y. Fakhri, A. Choukri, and N. Gherabi, "Comparative analysis of machine learning models for fake news detection in social media," IAES International Journal of Artificial Intelligence , vol. 14, no. 3, pp. 1951–1951, June 2025, doi: 10.11591/ijai.v14.i3.pp1951-1959.

[23] M. Al-Alshaqi, D. B. Rawat, and C. Liu, "Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, Natural Language Processing, and Machine Learning," Sensors , vol. 24, no. 18, pp. 6062–6062, Sept. 2024, doi: 10.3390/s24186062.

[24] R. K. Ayyasamy et al. , "A hybrid deep learning framework for fake news detection using LSTM-CGPNN and metaheuristic optimization," Scientific Reports , vol. 15, no. 1, pp. 41522–41522, Nov. 2025, doi: 10.1038/s41598-025-25311-x.