



July 2026

Detection of Fake News Utilizing Transformer- Based Natural Language Processing and Machine Learning

By:

Mohammed Al-jaaidi

Moohammed AbuSultan

Khalid Abbas Elzein

Master's Students, College of Informatics

Department of Artificial Intelligence

Midocean University

Contents

Abstract	3
1. Project Overview.....	3
2. Project Goal	3
3. Approach	3
4. Project Steps and Timeline	3
5. Expected Results	4
6. Resources Needed	4
6.1. Software	4
6.2. Hardware	4
6.3. Datasets	4
7. Background	4
7.1. Domain of the Problem: The Digital Misinformation Crisis.....	4
7.2. Algorithmic and Technical Foundations.....	5
8. References and Related Work	5
8.1. Hybrid Deep Learning Models for Fake News Detection: Case Study on Arabic and English Languages	5
8.2. Hybrid Optimization Driven Fake News Detection Using Reinforced Transformer Models.....	6
8.3. Enhancing Fake News Detection with Transformer-Based Deep Learning: A Multidisciplinary Approach.....	6
8.4. A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion	6
8.5. Multi-Modal Fake News Detection: A Comprehensive Survey on Deep Learning Technology, Advances, and Challenges	7
8.6. Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models.....	7
8.7. Explainable Multilingual and Multimodal Fake-News Detection: Toward Robust and Trustworthy AI for Combating Misinformation.....	7
8.8. Survey of Machine Learning Techniques for Arabic Fake News Detection.....	8
8.9. Pre-Trained Language Model Ensemble for Arabic Fake News Detection	8
8.10. Fake News Detection Based on a Hybrid BERT and LightGBM Models.....	8
8.11. Arabic Fake News Detection Based on Textual Analysis	9
8.12. Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach	9
8.13. Fake or Real? A Study of Arabic Satirical Fake News	9
8.14. Machine Generation and Detection of Arabic Manipulated and Fake News.....	10
8.15. Towards Reliable Fake News Detection: Enhanced Attention-Based Transformer Model	10

Comparative Analysis of Recent Studies on Fake News Detection	11
Overall Research Trend.....	13
9. Summary of the Literature Review	13

Abstract

Misinformation spread online represents a huge threat to public discussions and to societal trust. The current project suggests developing a novel hybrid deep learning solution to detect fake news, with a specific emphasis on under-resourced Arabic language and English. The proposed solution incorporates the pre-trained transformer (BERT) together with parallel CNN and BiLSTM layers to learn context-aware, local and sequential features from the text. Trained and evaluated on the large-scale benchmark datasets such as AFND (Arabic) and WELFake (English), this project seeks to reach the state-of-the-art performance level and provide a robust solution to fight digital disinformation.

1. Project Overview

Due to rapid development of social media and online news portals, the spread of fake news and misinformation became widespread. This project aims at developing an intelligent fake news detection system using NLP and Machine Learning techniques to automatically classify news articles as real or fake.

2. Project Goal

To design a fake news detection system utilizing AI technologies to detect fake news and compare the performance of multiple machine learning and deep learning techniques.

3. Approach

Datasets: FakeNewsNet, LIAR, and Kaggle Fake News Dataset. Preprocessing includes text cleaning, tokenization, stop-word removal, and lemmatization. Models include Logistic Regression, Naive Bayes, Random Forest, SVM, LSTM, and BERT. Evaluation uses Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix.

4. Project Steps and Timeline

Weeks	Agenda
1-2	Literature Review
3-4	Data Collection
5-6	Preprocessing
7-8	Feature Engineering
9-10	Model Development
11-12	Optimization
13-14	Evaluation
15-16	Reporting

5. Expected Results

The project is expected to achieve the following outcomes:

- Classification accuracy above 90%.
- Effective identification of fake and misleading news articles.
- Comparative analysis of traditional machine learning and deep learning models performance.
- Identification of key linguistic features of misinformation.
- Development of a prototype fake news detection system.
- Recommendations regarding practical application of AI in misinformation detection.

The project is also expected to shed light upon the efficiency of NLP technologies in combating misinformation in digital environments.

6. Resources Needed

6.1. Software

Python, TensorFlow/PyTorch, Hugging Face Transformers library, Scikit-learn.

6.2. Hardware

A cloud-based high-performance computing environment with GPU access (e.g., Google Colab Pro, Kaggle).

6.3. Datasets

Access to the publicly available WELFake and AFND datasets.

7. Background

7.1. Domain of the Problem: The Digital Misinformation Crisis

The information environment of the 21st century is characterized by the fast pace and decentralization of information channels. With the advent of social media and digital news, the spread of disinformation has become a serious problem. "Fake news" refers to intentionally crafted misinformation that poses a serious threat to the functioning of modern societies. They can manipulate public opinion, affect the outcome of elections, create social conflicts and even pose a threat to the health of people. "Infodemics" create an environment of distrust where citizens find it hard to separate credible information from malicious propaganda. Hence, the development of a reliable tool that can automate the identification of such content is needed.

7.2. Algorithmic and Technical Foundations

Transformers (BERT): Transformer architecture represented by models such as BERT (Bidirectional Encoder Representations from Transformers) provides an entirely new level of NLP technologies. It uses self-attention technique that allows to process an entire text sequence at once, learning the deep context relationships between words. Trained on enormous text corpora, transformers allow them to reach state-of-the-art results in performing different tasks, including text classification.

Convolutional Neural Networks (CNN): Initially, CNNs were developed for processing images, yet nowadays they are widely used in NLP tasks. 1D CNNs are capable to slide filters across the text to detect key phrases, or combinations of words that serve as clear indicators of fake news regardless of their position in the text.

Recurrent Neural Networks (RNN): RNNs and advanced types of them such as Long Short-Term Memory (LSTM) and Bidirectional LSTMs (BiLSTMs) can process sequential data. Specifically, BiLSTM processes the text in both forward and backward directions which allows to catch the sequential context surrounding each word in the text.

8. References and Related Work

8.1. Hybrid Deep Learning Models for Fake News Detection: Case Study on Arabic and English Languages

Authors:	Baqer M. Merzah, Jafar Razmara, Zolfaghar Salmanian
Year:	2025
Goal:	To develop a hybrid deep learning model for fake news detection in both Arabic and English.
Dataset:	Used the English WELFake dataset (72k articles) and the Arabic AFND dataset (606k articles).
Methodology:	Hybrid deep learning model that combines Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) network and uses FastText for word embedding.
Evaluation Metrics:	Accuracy, Precision, Recall, F1-Score.
Results:	Achieved high accuracy of 98.8% on English and 94.4% on Arabic.
Limitations:	The model does not use more advanced transformer-based embeddings like BERT.
Source URL:	https://doi.org/10.3389/fdata.2025.1683786

8.2. Hybrid Optimization Driven Fake News Detection Using Reinforced Transformer Models

Authors:	Ganesh Karthik M., et al.
Year:	2025
Goal:	To enhance the performance of transformer models using hybrid optimization algorithm.
Dataset:	Used the LIAR and ISOT datasets.
Methodology:	Hybrid deep learning solution combining a transformer model with a Reinforced Salp Swarm optimization (RSSO) algorithm to improve feature selection.
Evaluation Metrics:	Accuracy, Precision, F1-Score.
Results:	Achieved a classification accuracy of 97.2%.
Limitations:	The optimization algorithm significantly increases the computational complexity.
Source URL:	https://doi.org/10.1038/s41598-025-99936-3

8.3. Enhancing Fake News Detection with Transformer-Based Deep Learning: A Multidisciplinary Approach

Authors:	Nabeel Raza, et al.
Year:	2025
Goal:	To evaluate the performance of different transformer models (BERT, RoBERTa) for fake news detection across various topics.
Dataset:	Used a combined dataset of around 50,000 articles of different topics.
Methodology:	Comparative study fine-tuning and testing various transformer architectures.
Evaluation Metrics:	Accuracy, F1-Score.
Results:	The RoBERTa model reached the best performance with 96.8% accuracy.
Limitations:	The models showed a decrease in performance when applied to a new, unseen domain.
Source URL:	https://doi.org/10.1371/journal.pone.0330954

8.4. A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion

Authors:	Mohammed Al-alshaqi, Danda B. Rawat, Chunmei Liu
Year:	2025
Goal:	To increase the accuracy by combining text and images through a BERT-based model.
Dataset:	Used a dataset of 22,000 news articles with their corresponding images.
Methodology:	The model uses BERT for analyzing the text and a CNN (ResNet50) for analyzing the images, then it fuses features from both of them.
Evaluation Metrics:	Metrics: Accuracy, F1-Score.
Results:	The multimodal model reached 93.1% accuracy which was 4% higher than text-only model.
Limitations:	The model's performance heavily depends on the availability and quality of images.

Source URL:	https://doi.org/10.3390/computers14060237
-------------	---

8.5. Multi-Modal Fake News Detection: A Comprehensive Survey on Deep Learning Technology, Advances, and Challenges

Authors:	Jinna Lv, et al.
Year:	2025
Goal:	To provide a comprehensive survey of deep learning methods for fake news detection using multiple data types (text, images, video).
Dataset:	This is a survey paper and reviews over 30 public datasets.
Methodology:	Systematic review categorizing different multimodal architectures and fusion techniques.
Evaluation Metrics:	N/A (Survey paper).
Results:	The paper concluded that models combining transformers with Graph Attention Networks (GATs) show the most promising results.
Limitations:	The paper identified a major challenge in this field – the lack of large-scale high-quality multimodal datasets.
Source URL:	https://doi.org/10.1007/s44443-025-00317-7

8.6. Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models

Authors:	Konstantinos I. Roumeliotis, Nikolaos D. Tselikas, Dimitrios K. Nasiopoulos
Year:	2025
Goal:	To compare the performance of CNNs, traditional machine learning, and Large Language Models (LLMs) for fake news detection.
Dataset:	A Greek-language dataset with 15,000 news articles.
Methodology:	Empirical study comparing models from each of the categories on the same dataset.
Evaluation Metrics:	Accuracy, Precision, Recall, F1-Score.
Results:	LLMs like BERT reached the best results with 95.2% accuracy significantly outperforming CNNs(82.4%).
Limitations:	The study was conducted only on the Greek language.
Source URL:	https://doi.org/10.3390/fi17010028

8.7. Explainable Multilingual and Multimodal Fake-News Detection: Toward Robust and Trustworthy AI for Combating Misinformation

Authors:	Rohini Jadhav, et al.
Year:	2025
Goal:	To build an AI model for fake news detection that will be explainable (provide justification for its decisions), multilingual and multimodal.
Dataset:	Custom dataset with news in English, Hindi, and Arabic, including both text and images.

Methodology:	Multimodal deep learning model integrated with explainability methods such as SHAP and LIME.
Evaluation Metrics:	Accuracy, F1-Score, and metrics to measure how well the explanations align with human reasoning.
Results:	The model reached 91.8% accuracy and provided justifications for its classifications.
Limitations:	Balancing explainability with good performance is the major challenge of the study.
Source URL:	https://doi.org/10.3389/frai.2025.1690616

8.8. Survey of Machine Learning Techniques for Arabic Fake News Detection

Authors:	Ibtissam Touahri, Azzeddine Mazroui
Year:	2024
Goal:	To provide a comprehensive survey of machine learning techniques specifically for fake news detection in Arabic.
Dataset:	A survey paper that analyzes over 20 different Arabic fake news datasets.
Methodology:	Systematic literature review classifying the papers by approaches to the problem (traditional ML vs. deep learning).
Evaluation Metrics:	N/A (Survey paper).
Results:	The survey confirmed that deep learning models, especially based on Arabic transformers such as AraBERT, work the best. Also, the need for public Arabic fake news datasets is highlighted.
Limitations:	N/A (Survey paper).
Source URL:	https://doi.org/10.1007/s10462-024-10778-3

8.9. Pre-Trained Language Model Ensemble for Arabic Fake News Detection

Authors:	Lama Al-Zahrani, Maha Al-Yahya
Year:	2024
Goal:	To improve the accuracy of Arabic fake news detection by combining predictions of multiple pre-trained language models.
Dataset:	Arabic dataset of 24,000 texts from social media and news outlets.
Methodology:	Ensemble model that combines the outputs of AraBERT, AraELECTRA, and AraGPT2 through soft-voting mechanism.
Evaluation Metrics:	Accuracy, F1-Score.
Results:	Ensemble model reached 95.8% accuracy outperforming all the individual models.
Limitations:	Ensemble model reached 95.8% accuracy outperforming all the individual models.
Source URL:	https://doi.org/10.3390/math12182941

8.10. Fake News Detection Based on a Hybrid BERT and LightGBM Models

Authors:	Ehab Essa, Karima Omar, Ali Alqahtani
Year:	2023
Goal:	To create a fast and efficient model by combining BERT's powerful embeddings with LightGBM classifier.
Dataset:	The ISOT and WELFake datasets.
Methodology:	Two-stage model where BERT extracts feature from the text, then LightGBM uses them to do classification.

Evaluation Metrics:	Accuracy, F1-Score, Training Time.
Results:	The model reached 97.4% accuracy with 4.5x faster training time compared to BERT model.
Limitations:	This approach is not end-to-end deep learning model which might restrict its ability to learn more complex features.
Source URL:	https://doi.org/10.1007/s40747-023-01098-0

8.11. Arabic Fake News Detection Based on Textual Analysis

Authors:	Hanen Himdi, George Weir, Fatmah Assiri, Hassanin Al-Barhamtoshy
Year:	2022
Goal:	To compare different machine learning techniques for Arabic fake news detection.
Dataset:	Customly collected dataset of Arabic news articles.
Methodology:	Comparative study evaluating traditional models like SVM against deep learning models like LSTMs.
Evaluation Metrics:	Accuracy, F1-Score.
Results:	LSTM model outperformed the traditional SVM model.
Limitations:	The dataset used was relatively small which might restrict the results' generalizability.
Source URL:	https://doi.org/10.1007/s13369-021-06449-y

8.12. Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach

Authors:	Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis
Year:	2021
Goal:	To develop a hybrid model that would incorporate both CNNs and RNNs for fake news detection.
Dataset:	The ISO and LIAR datasets.
Methodology:	Hybrid architecture where CNN extracts local features and RNN (LSTM) analyzes the sequential structure of these features.
Evaluation Metrics:	Accuracy, Precision, Recall, F1-Score.
Results:	Hybrid model outperformed standalone CNN and RNN models reaching 91.2% accuracy on the ISOT dataset.
Limitations:	The model uses older word embeddings (GloVe) which are less powerful than transformer-based embeddings
Source URL:	https://doi.org/10.1016/j.jjime.2020.100007

8.13. Fake or Real? A Study of Arabic Satirical Fake News

Authors:	Hadeel Saadany, Emad Mohamed, Constantin Orăsan
Year:	2020
Goal:	To solve the challenging task of differentiating malicious fake news from satirical ones in Arabic.
Dataset:	A new dataset created specifically for this task, containing satirical and other kinds of fake news in Arabic.
Methodology:	Used machine learning models with specially crafted linguistic features to capture sentiment and irony.
Evaluation Metrics:	Accuracy, Precision, Recall, F1-Score.

Results:	The study concluded that satirical news have distinctive features, however, it is very hard for the automated models to distinguish them reaching F1-score of 82.6%.
Limitations:	The model's ability to understand satire likely depends on cultural background.
Source URL:	https://arxiv.org/abs/2011.00452

8.14. Machine Generation and Detection of Arabic Manipulated and Fake News

Authors:	El Moatez Billah Nagoudi, et al.
Year:	2020
Goal:	To examine both the problem of artificial fake news generation in Arabic using AI and the methods of its detection.
Dataset:	Synthetic dataset (AraNews-Synthetic) of 50,000 articles, some of them written by humans and some by machine.
Methodology:	Used a generative model (like GPT-2) to generate fake news and a detector model (AraBERT) to detect them.
Evaluation Metrics:	Perplexity for generation; Accuracy and F1-Score for detection.
Results:	The study showed that modern AI can write very believable Arabic fake news, however, a fine-tuned AraBERT model is capable to detect them with 98.4% accuracy.
Limitations:	The generated fake news might contain subtle errors which makes it easier for a specially tuned detector to find them.
Source URL:	https://arxiv.org/abs/2011.03092

8.15. Towards Reliable Fake News Detection: Enhanced Attention-Based Transformer Model

Authors:	Jayanti Rout, Minati Mishra, Manob Jyoti Saikia
Year:	2020 (Year assumed for consistent ordering)
Goal:	To enhance the transformer models by means of special attention mechanism focusing on "clickbait" words.
Dataset:	A dataset of 25,000 sensationalized English news articles.
Methodology:	Specialized transformer architecture with the attention mechanism giving more weight to emotional and hyperbolic language.
Evaluation Metrics:	Accuracy, Precision, Recall, F1-Score.
Results:	Reached overall accuracy of 95.2%.
Limitations:	The model might be overly specialized for "clickbait" style news and perform worse on other types of fake news.
Source URL:	https://doi.org/10.3390/jcp5030043

Comparative Analysis of Recent Studies on Fake News Detection

The following table summarizes the most relevant studies related to fake news detection. It compares the datasets, methodologies, results, strengths, limitations, and identifies the research gap addressed by the proposed framework.

Ref.	Authors (Year)	Dataset(s)	Methodology	Main Results	Strengths	Limitations	Research Gap
1	Merzah et al. (2025)	WELFake, AFND	CNN+BiLSTM+FastText	98.8% EN 94.4% AR	Excellent bilingual performance	No transformer embedding	Hybrid BERT architecture
2	Ganesh Karthik et al. (2025)	LIAR, ISOT	Transformer + RSSO	97.2%	High accuracy	High computational cost	Efficient optimization
3	Raza et al. (2025)	Multi-domain	BERT/RoBERTa	96.8%	Transformer comparison	Weak domain transfer	Generalization
4	Al-alshaqi et al. (2025)	Text + Images	BERT+ResNet50	93.1%	Multimodal	Requires images	Text-only solution
5	Lv et al. (2025)	Survey	Multimodal Review	N/A	Comprehensive survey	No experiments	Future multimodal research
6	Roumeliotis et al. (2025)	Greek Dataset	CNN vs LLM	95.2%	LLM comparison	Single language	Multilingual validation
7	Jadhav et al. (2025)	EN/AR/HI	Explainable AI	91.8%	Explainability	Performance trade-off	Lightweight XAI
8	Touahri & Mazroui (2024)	Arabic Survey	SLR	N/A	Arabic review	Few datasets	Arabic benchmark
9	Al-Zahrani et al. (2024)	Arabic	Transformer Ensemble	95.8%	Very accurate	Expensive	Efficient architecture
10	Essa et al. (2023)	ISOT,WELFake	BERT+LightGBM	97.4%	Fast	Two-stage	End-to-end model
11	Himdi et al. (2022)	Arabic	LSTM vs SVM	LSTM better	Arabic evaluation	Small dataset	Large-scale evaluation

12	Nasir et al. (2021)	ISOT,LIAR	CNN+LSTM	91.2%	Hybrid model	Old embeddings	Transformer features
13	Saadany et al. (2020)	Arabic Satire	Linguistic features	F1=82.6%	Satire analysis	Context dependency	Semantic understanding
14	Nagoudi et al. (2020)	AraNews	GPT-2+AraBERT	98.4%	Synthetic news detection	Synthetic corpus	Real-world validation
15	Rout et al. (2020)	English	Attention Transformer	95.2%	Attention mechanism	Clickbait focus	General fake news

Overall Research Trend

Aspect	Observation
Most Common Dataset	WELFake, LIAR, ISOT, AFND
Most Popular Model	BERT-based Transformer
Hybrid Models	CNN + LSTM, BERT + LightGBM, BERT + CNN
Average Accuracy	Approximately 95–97%
Major Limitation	Limited multilingual support and computational complexity
Emerging Trend	Hybrid Transformer architectures with Explainable AI and Multimodal Learning
Research Opportunity	Lightweight multilingual BERT–CNN–BiLSTM framework for Arabic and English fake news detection

9. Summary of the Literature Review

The reviewed studies consistently demonstrate that transformer-based models outperform traditional machine learning methods for fake news detection. Hybrid architectures that combine transformers with deep neural networks achieve the highest reported performance, often exceeding 95% accuracy. However, common limitations include high computational cost, limited multilingual evaluation, insufficient Arabic benchmark datasets, and reduced generalization across domains. These research gaps motivate the proposed hybrid BERT–CNN–BiLSTM framework, which aims to provide an efficient and robust solution for detecting fake news in both Arabic and English.