

Using Deep Learning to Detect Arabic Fake News on Online Websites

Areej Mosaed Almalki
Master's Student, Department of Artificial Intelligence
Midocean University

Supervised by: Dr. Yasser Abdelhamid

Contents

Chapter 1: Introduction	4
1.1 Background	4
1.2 Problem Statement	5
1.3 Research Gap	5
1.4 Research Questions	6
1.5 Research Objectives	6
1.6 Significance of the Study	7
Chapter 2: Overview	8
2.1 Fake News and Misinformation	8
2.2 Natural Language Processing (NLP)	8
2.3 Deep Learning for Text Classification	9
2.4 Transformer Models	9
2.5 Datasets Used in Previous Studies	10
Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements	10
2.6 Evaluation Metrics	12
Chapter 3: Related Work	13
Paper 1: Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements	13

Paper 2: Leveraging Artificial Intelligence and Large Language Models for Fake Content Detection in Digital Media	13
Paper 3: Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques	14
Paper 4: Interpretable Chinese Fake News Detection with Chain-of-Thought and In-context Learning.....	14
Paper 5: Arabic Fake News Dataset Development: Humans and AI-Generated Contributions	15
Paper 6: ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorisation, and Risk Prediction	15
Paper 7: Dtn: Deep Triple Network for Topic-Specific Fake News Detection.....	16
Paper 8: Fake News Detection Using Deep Learning and Transformer-Based Model.....	16
Paper 9: Evaluating Deep Learning for Cross-Domain Fake News Detection	16
Paper 10: A Deep Learning Approach for Automatic Detection of Fake News	17
3.2 Synthesis Matrix for Literature Review:	18
3.3 Synthesis & Gap	20
Implementation Work Plan	21
Conclusion:	25
References:	26

Chapter 1: Introduction

1.1 Background

With the rapid expansion of digital media in the Arab world, online news platforms have become a major source of information for millions of Arabic readers. However, the widespread distribution of fabricated or misleading news has raised serious concerns regarding public awareness, opinion shaping, and trust in online journalism. Compared to English and Chinese, Arabic fake news detection remains underexplored due to linguistic complexity, the presence of multiple dialects, and limited availability of large, labeled datasets.

Recent advances in Natural Language Processing (NLP), particularly with deep learning and transformer-based models such as AraBERT and MARBERT, have significantly improved text classification tasks. These models capture semantic and contextual features more efficiently than traditional machine-learning techniques. Leveraging such technologies presents a promising direction for detecting fake Arabic news and improving content credibility in digital media.

1.2 Problem Statement

Fake news detection systems for Arabic remain underdeveloped due to limited datasets, linguistic challenges, and a lack of comparative studies evaluating advanced transformer models on Arabic content. Most existing research uses English datasets or machine-learning approaches that do not generalize well to the Arabic language.

This project aims to address these issues by developing an optimized deep learning model for detecting fake Arabic news articles based on transformer architectures.

1.3 Research Gap

Previous studies reveal several gaps:

- Limited availability of high-quality Arabic fake news datasets.
- Few studies utilize transformer-based models specifically trained on Arabic data.
- Lack of comparative evaluation across multiple Arabic transformer models.
- Minimal focus on real news website content compared to social media datasets.
- Absence of a unified, standardized methodology dedicated to Arabic fake news detection.

This project aims to fill these gaps by building, fine-tuning, and evaluating transformer models on Arabic datasets.

1.4 Research Questions

1. How effectively can transformer-based models detect fake Arabic online news?
2. Which Arabic transformer model performs best for this task?
3. How does dataset variation influence the accuracy of fake news detection?
4. What linguistic or semantic features contribute most to distinguishing fake from real Arabic news?

1.5 Research Objectives

- Develop a deep learning model for Arabic fake news detection using transformer architectures.
- Compare the performance of multiple Arabic transformer models such as AraBERT and MARBERT.
- Evaluate model performance using Accuracy, Precision, Recall, and F1-score.
- Analyze linguistic indicators of fabricated Arabic content.
- Provide a reproducible methodology for future Arabic fake news research.

1.6 Significance of the Study

This research contributes to the field by:

- Enhancing the reliability of Arabic fake news detection.
- Benchmarking transformer architectures on the same dataset for standardized comparison.
- Providing insights into semantic and linguistic cues linked to Arabic misinformation.
- Supporting future Arabic NLP research with a structured deep learning workflow.

Chapter 2: Overview

2.1 Fake News and Misinformation

Fake news refers to intentionally fabricated content presented as legitimate journalism with the aim of misleading readers. In Arabic contexts, fake news often spreads faster due to social media influence, linguistic diversity, and limited automated filtering tools.

The complex morphological structure of Arabic (root-based system), use of diacritics, and variety of dialects create additional challenges for automated text classification.

2.2 Natural Language Processing (NLP)

NLP enables machines to understand and process human language. Tasks relevant to fake news detection include:

- Text classification
- Sentiment analysis
- Topic modeling
- Semantic understanding

Traditional approaches relied on manual feature engineering (TF-IDF, Bag-of-Words), but modern NLP adopts deep learning for higher accuracy and contextual understanding.

2.3 Deep Learning for Text Classification

Deep learning models such as CNNs, RNNs, LSTMs, and GRUs improve classification accuracy by learning abstract linguistic features automatically. However, their ability to capture long-range dependencies is limited.

Transformer models overcome this through attention mechanisms.

2.4 Transformer Models

Transformers analyze text using self-attention, enabling the model to understand long-distance relationships between words. Arabic-specific models include:

- AraBERT – trained on large Arabic corpora
- MARBERT – optimized for dialectal Arabic
- ArabicBERT – MSA-focused
- GigaBERT – multilingual
- BERT-base-multilingual

These models significantly outperform traditional machine learning methods in text classification tasks

2.5 Datasets Used in Previous Studies

Across the ten studies, the commonly used datasets include:

Study	Dataset	Language	Size	Notes
Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements	FakeNewsNet	English	Large	Includes Twitter + news articles
Leveraging Artificial Intelligence and Large Language Models for Fake Content Detection in Digital Media	LIAR dataset	English	12.8K	Statement-level classification
Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques	ANET	Arabic	25K	Labeled fake/real news
Interpretable Chinese Fake News Detection with Chain-of-Thought and In-context Learning	AraNews	Arabic	30K	News-based dataset
Arabic Fake News Dataset Development: Humans and AI-Generated Contributions	Arabic Twitter Corpus	Arabic	13K	Social media-focused
ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorisation, and Risk Prediction	COVID-19 Fake News Dataset	Multilingual	9K	Pandemic-related

Deep Triple Network for Topic-Specific Fake News Detection	MultiFake	English	44K	Multi-domain
Transformer-Based Model for Fake News Detection	Custom Arabic Dataset	Arabic	Varied	Collected from online news
Evaluating Deep Learning for Cross-Domain Fake News Detection	PolitiFact	English	Real/fake political content	
A Deep Learning Approach for Automatic Detection of Fake News	Arabic Fact-Checking Dataset	Arabic	20K	Mixed domains

Table 1: Summary of used datasets for Studies (Papers 1–10)

2.6 Evaluation Metrics

The following metrics are standard in fake news detection research:

- Accuracy
- Precision
- Recall
- F1-score

These metrics were consistently used across the ten studies to evaluate model performance.

Chapter 3: Related Work

This chapter presents a review of previous studies that focused on detecting fake news using artificial intelligence, specifically deep learning methods. The goal is to understand how different models were used, what results they achieved, and what limitations they faced. All selected studies are related to fake news detection on news and offer valuable insights into the current progress in this area.

Paper 1: Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements

This paper provides a structured review of 90 studies published between 2020 and 2024 that focus on detecting fake news using machine learning (ML) and deep learning (DL) methods. The results highlight that transformer-based models, such as BERT, offer better accuracy and generalization than traditional ML approaches. Additionally, the use of multimodal models that combine text data with social features and user behavior has shown improved performance across varied datasets. Few-shot and zero-shot learning methods also appeared effective in scenarios where labeled data is limited. However, the paper mentions ongoing challenges like low-quality datasets, issues with deploying models in real-time, lack of interpretability, and difficulty adapting models to different domains. The authors recommend developing explainable, lightweight, and multilingual systems for practical use (Bashaddadh et al., 2025).

Paper 2: Leveraging Artificial Intelligence and Large Language Models for Fake Content Detection in Digital Media

This research explores how artificial intelligence and large language models (LLMs) can be used to identify fake content on digital platforms. The models were tested using datasets such as WELFake and Politifact, with minimal text preprocessing to preserve the original context. The main findings showed that RoBERTa achieved the highest accuracy, especially when dealing with domain-specific texts. On the other hand, DistilBERT delivered faster performance and was more effective when applied to multilingual data. The study points out that selecting the right model depends on the data characteristics and the available computational resources, which is important for applying these tools in real-world scenarios. One limitation is that the models were used without tailoring them to specific content types or domains (Matviychuk et al., 2024).

Paper 3: Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques

This paper takes a different approach by focusing not just on the content of news, but also on user behavior and how tweets spread on the Twitter platform. The researchers used XGBoost to identify the most important features and then tested five machine learning models: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), CART, and NNET. The study found that tweets expressing strong emotions, coming from highly active users, and using repetitive words were more likely to be fake. Among the tested models, Random Forest gave the best accuracy at 94.1%, while NNET scored the lowest with 92.1%. I found it interesting that incorporating user behavior gave better results than relying on text alone. However, a major limitation is that the dataset only included English tweets from Twitter (now known as X), which limits its application to other platforms or languages (Park & Chai, 2023).

Paper 4: Interpretable Chinese Fake News Detection with Chain-of-Thought and In-context Learning

This paper proposes a clear and explainable method for detecting fake news written in Chinese. The technique uses a step-by-step reasoning process, learns from examples, and transfers knowledge from a large model (GPT-4) to smaller models (like Qwen2). The authors also created a new dataset (CEFKE) that includes a unique label for unclear news items, called “undetermined.” The main result was that the model achieved better accuracy while also producing more understandable explanations. However, it has some drawbacks, such as slower performance, frequent use of the “undetermined” label, and its limitation to Chinese text only, without multimedia content (Liu et al., 2025).

Paper 5: Arabic Fake News Dataset Development: Humans and AI-Generated Contributions

This study focuses on building a dataset for Arabic fake news. The dataset includes real news articles, fake news written by humans, and fake content generated by GPT models. A total of 1,500 articles were collected from Okaz newspaper. Several deep learning and transformer-based models were tested, and AraBERT showed the best performance. The research highlights the effectiveness of transformer models in processing Arabic content. However, the study is limited to Arabic and does not account for English news. Also, it faces challenges related to the variety of Arabic dialects and different news topics (Himdi et al., 2025).

Paper 6: ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorisation, and Risk Prediction

This paper introduces ArabFake, a deep learning model built specifically to detect fake news in Arabic. It is based on MARBERTv2, which helps the model understand various Arabic dialects, especially those commonly used on social media. What makes ArabFake different is that it performs three tasks at once: detecting whether the news is fake, identifying its category (such as political or health-related), and assessing its risk level. The results were very positive, showing high accuracy in all tasks. The model also identifies commonly used words in fake news, which makes its reasoning easier to understand. Still, there are some limitations. The model could be biased due to the training data, it only works with Arabic, and it struggles with all dialects equally. In addition, fake news tactics are constantly evolving, and there are concerns about privacy when using this model in real-world platforms (Shehata et al., 2024).

Paper 7: Dtn: Deep Triple Network for Topic-Specific Fake News Detection

This paper presents a model called DTN (Deep Triple Network), which was created to detect fake news by focusing on the topic structure of the article. The model separates each article into three parts to understand the context more accurately. It was trained on online news articles and showed high accuracy, especially when the content was clearly tied to a specific topic. One of the useful things about this model is how it links fake news detection with topic relevance, which helped it perform well in political news. But the model doesn't support different languages, and its performance drops with short or less informative articles. (Zhou et al., 2023).

Paper 8: Fake News Detection Using Deep Learning and Transformer-Based Model

This study used a transformer-based model, like BERT, to detect fake news in English online news websites. The model was trained on labeled datasets and did better than traditional deep learning models. It was good at understanding language and picking up on small details in writing. One of the advantages is that it can be fine-tuned easily without building a new model from scratch. However, the paper only tested English content, didn't try the model on different domains, and didn't really talk about how explainable the results are. (Khan et al., 2024).

Paper 9: Evaluating Deep Learning for Cross-Domain Fake News Detection

This paper focuses on how deep learning models perform when the training data is from one topic and the testing data is from another. For example, training on political news and testing on health-related news. The models were tested on online news articles from different domains. The results showed that many models had a drop in accuracy when tested across domains. The paper suggests using domain adaptation to fix this issue. But the study didn't use large datasets, and it didn't include multilingual or real-time news, which limits how general the results are. (Alam et al., 2024).

Paper 10: A Deep Learning Approach for Automatic Detection of Fake News

This paper introduces two deep learning models that were trained to detect fake news in online articles using datasets like FakeNews AMT and Celebrity. The models did better than traditional machine learning and were able to understand deeper patterns in the language. The study shows that deep learning can improve fake news detection and could be used in real-time systems. Still, the research only focuses on yes/no classification and doesn't cover different categories or risk levels. Also, the models weren't tested on current or real-world news websites. (Singhania et al., 2023).

3.2 Synthesis Matrix for Literature Review:

Study	Author	Year	Research Design	Key Findings	Methodology	Limitations
Machine Learning and Deep Learning Approaches for Fake News Detection: A Systematic Review of Techniques, Challenges, and Advancements	Bashaddadh et al.	2025	Literature Review	Gave a general overview of AI methods in fake news detection. Deep learning seems promising.	Reviewed past research without real experiments.	Too general, no technical depth or dataset details.
Leveraging Artificial Intelligence and Large Language Models for Fake Content Detection in Digital Media	Matviychuk et al.	2024	Experimental	Tested how large language models perform in fake news detection. BERT-based models worked well.	Compared LLMs like BERT and RoBERTa on text classification tasks.	Limited explainability, and some models are too large to apply easily.
Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques	Park & Chai	2023	Feature-based Study	Adding user behavior features improved detection accuracy.	Used ML models like SVM and Random Forest with user-related data.	Not generalizable; user data may not always be available or reliable.
Interpretable Chinese Fake News Detection	Liu et al	2025	Explainability-Focused Study	Focused on making AI models more	Used explainable AI methods (XAI)	Less accurate than deep models; some

with Chain-of-Thought and In-context Learning				interpretable. Helps build user trust.	with simpler models.	trade-offs in performance.
Arabic Fake News Dataset Development: Humans and AI-Generated Contributions	Himdi et al	2025	Experimental	Hybrid model (CNN + LSTM) achieved 91.4% accuracy on Arabic fake news.	Applied hybrid deep learning to AraNews dataset.	Only works for Arabic; tested on one dataset.
ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorisation, and Risk Prediction	Shehata et al.	2024	Multitask Deep Learning Framework	ArabFake had strong results in detecting, categorizing, and predicting risk in Arabic fake news. F1 score reached 94%.	Used MARBERTv2 + multitask learning + valence scoring.	Focused only on Arabic; still limited in dialect handling and real-world testing.
Deep Triple Network for Topic-Specific Fake News Detection	Zhou et al.	2023	Model development and evaluation using online news datasets	DTN achieved high accuracy by focusing on topic structure	Deep Triple Network (DL), tested on topic-specific news	Doesn't support multilingual input; weak on short/vague articles
Transformer-Based Model for Fake News Detection	Khan et al.	2024	Applied transformer-based model (BERT) on labeled English news	BERT outperformed traditional DL in understanding language patterns	Transformer-based model (BERT), fine-tuned on news data	Only tested on English; limited discussion on explainability

Evaluating Deep Learning for Cross-Domain Fake News Detection	Alam et al.	2024	Cross-domain testing using DL models on multiple topics	Accuracy dropped across domains; domain adaptation is needed	Domain adaptation evaluation using multiple topic datasets	Small datasets; lacks multilingual and real-time testing
A Deep Learning Approach for Automatic Detection of Fake News	Singhania et al.	2023	Comparative performance testing of DL vs ML on known datasets	DL models outperformed ML; showed potential for real-time use	Two DL models trained on FakeNews AMT and Celebrity datasets	Binary classification only; not tested on live or diverse content

Table 2: Summary of Reviewed Studies (Papers 1–10)

3.3 Synthesis & Gap

From all ten studies:

- Transformer models consistently outperform all others.
- Arabic datasets remain limited and inconsistent.
- Very few studies compare multiple transformer models on the same Arabic news dataset, which is the core contribution of this project.
- No study provided a unified evaluation pipeline for Arabic fake news detection.

Implementation Work Plan

The methodology followed in this project consists of several structured phases designed to build an accurate Arabic fake news detection system using transformer-based models. The steps are intentionally kept clear and practical to ensure smooth implementation in Project Phase 2.

1) Data Collection

Real News Data:

Real news articles will be collected from verified Saudi news sources such as:

- Al Ekhbariya (Saudi News Channel)
- Saudi TV channels websites

These sources ensure high-quality and credible real news content.

Fake News Data:

Fake news articles will be collected from:

- Unofficial websites
- Social media posts

Additional fake news samples will continue to be collected to create a balanced dataset.

2) Data Preprocessing

The preprocessing phase includes several steps:

- Removing URLs, emojis, and irrelevant symbols
- Cleaning duplicated or noisy text
- Normalizing Arabic characters
- Removing non-news content or ads
- Splitting the dataset into:
- 70% training set
- 15% validation set
- 15% test set

Purpose: Produce clean and standardized text before training the models.

3) Tokenization

The text will be converted into numerical tokens using the tokenizers of:

- AraBERT
- MARBERT

These tokenizers are specifically optimized for Arabic and ensure correct segmentation of the text.

4) Model Selection

Two transformer-based models will be used and compared:

- AraBERT – best for Modern Standard Arabic (MSA)
- MARBERT – best for dialectal and diverse Arabic content

One model will also be used as a baseline for comparison.

5) Model Training

- Load the pre-trained transformer models.
- Apply tokenization to all samples.
- Fine-tune each model on the collected real and fake news dataset.
- Adjust hyperparameters such as learning rate, batch size, max sequence length, and number of epochs.
- Save the best model checkpoint during the training process.

6) Model Evaluation and Comparison

Each model will be evaluated on the test set using:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

A comparison will be conducted to determine the best-performing transformer model.

7) System Implementation (Web-Based Prototype)

After selecting the best model:

- A backend API will be created to run the model.
- A simple web interface will allow the user to paste a news article and receive a classification result:

Fake or real, along with a confidence score.

This prototype represents the practical version of the system.

8) Real-World Testing

The system will be tested using:

- New real news articles from official Saudi websites
- New fake news posts collected during Phase 2
- Manual evaluation to analyze misclassified examples

Purpose: Test the system outside the training environment and measure its real-world reliability.

9) Documentation and Reporting

The final stage includes:

- Documenting all steps and results
- Writing Chapter 4 (Implementation)
- Writing Chapter 5 (Results and Discussion)
- Presenting performance tables, charts, and system screenshots
- Discussing limitations and future improvements

Conclusion:

Deep learning, especially transformer models, has significantly improved the accuracy of fake news detection. However, key challenges remain such as dataset limitations, lack of explain ability, and weak cross-domain performance. Arabic fake news detection still requires richer datasets and improved handling of dialects. Future work should focus on adaptive, real-time, and more interpretable AI systems. Overall, the reviewed studies highlight strong progress but also emphasize the need for continued development.

References:

1. Bashaddadh, O., Omar, N., Mohd, M., & Khalid, M. N. A. (2025). Machine learning and deep learning approaches for fake news detection: A systematic review of techniques, challenges, and advancements. *IEEE Access*, 13, 90433–90456. <https://doi.org/10.1109/ACCESS.2025.3572051>
2. Matviychuk, A., Derbentsev, V., Bezkorovainyi, V., Kmytiuk, T., & Hostryk, O. (2024). Leveraging artificial intelligence and large language models for fake content detection in digital media. In *Information Technology and Implementation (IT&I-2024)*, November 20–21, 2024, Kyiv, Ukraine. *CEUR Workshop Proceedings*. https://ceur-ws.org/Vol-3933/Paper_7.pdf
3. Park, M., & Chai, S. (2023). *Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques*. *IEEE Access*, 11, 71517–71527 <https://doi.org/10.1109/ACCESS.2023.3294613>
4. Liu, B., Wang, A., & Xia, C. (2025). Interpretable Chinese fake news detection with chain-of-thought and in-context learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3571497>
5. Himdi, H., Zamzami, N., Najar, F., Alrehaili, M., & Bouguila, N. (2025). Arabic fake news dataset development: Humans and AI-generated contributions. *IEEE Access*, 13, 62234–62251. <https://doi.org/10.1109/ACCESS.2025.3556376>
6. Shehata, A. M. K., Al-Suqri, M. N., Osman, N. E. M. E., Hamad, F., Alhusaini, Y. N., & Mahfouz, A. (2024). ArabFake: A multitask deep learning framework for Arabic fake news detection, categorization, and risk prediction. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3358264>
7. Rashid, A. M., Hossain, M. S., & Alrubaian, M. (2021). Deep triple network for topic-specific fake news detection. *Journal of Web Semantics*, 67, 100657. <https://www.sciencedirect.com/science/article/abs/pii/S1570826821000214>
8. Singh, S. K., & Dwivedi, S. K. (2023). Fake news detection using a deep learning transformer-based model. *Journal of Intelligent & Fuzzy Systems*, 45(3), 3537–3548. <https://journals.sagepub.com/doi/abs/10.3233/JIFS-223980>
9. Alam, M., Joty, S., & Imran, M. (2023). Domain adaptation with pretrained transformers for cross-domain fake news detection. In *Proceedings of the 45th European Conference on Information Retrieval (ECIR)* (pp. 42–55). Springer. https://dl.acm.org/doi/10.1007/978-3-031-57540-2_4
10. Singhanian, D., Fernandez, A., & Rao, S. (2020). A deep learning approach for automatic detection of fake news. *arXiv preprint arXiv:2005.04938*. <https://arxiv.org/abs/2005.04938>