

Course Title: Research Methodology in Computer Science

Courses code: AI800

Weighting: 20%

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 28 June 2025

Hand -in Date: 12 July 2025

Hand -in Method: Online Submission via Moodle platform

Assignment Report: Using Artificial Intelligence to Detect and Handle Fake News: Gaps, Challenges, and Future Direction

Table of Contents:

- Summary Research papers
- Critical Analysis
- Connection
- Conflict between the findings
- Synthesis Matrix for Literature Review
- References

Summary Research papers:

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).

Critical Analysis:

After reading all six papers, each one brings something useful but also has some gaps. Some of the studies were very focused on technical results like accuracy but didn't really talk about real-life challenges, like how biased the data might be or how these models could affect people. For example, Papers 1 and 2 gave good overviews but felt a bit general. Papers 3 and 4 were more about the user side, which I liked, but they didn't seem easy to apply on a big scale. What stood out to me the most were Papers 5 and 6 because they talked about Arabic fake news, which isn't discussed enough in most research. Still, the fact that they only focused on Arabic makes it hard to apply their findings to other languages or regions. In the end, I think all the papers are helpful, but most of them need more work when it comes to explaining how their models actually work in the real world.

In addition to the general observations, I also noticed a few technical aspects that stood out across the six studies. First, transformer-based models like BERT, RoBERTa, and MARBERTv2 clearly outperform traditional machine learning methods such as SVM or Random Forest in most cases. This is mainly because transformers can understand the context of words using attention mechanisms, while traditional models rely on simpler, sparse representations. However, these deep models often require more data and computational resources, which can be a limitation for real-time systems.

For example, in Paper 2, RoBERTa showed high accuracy on domain-specific tasks thanks to dynamic masking and a larger training corpus. Still, the model is resource-intensive compared to smaller models like DistilBERT, which trade off a bit of accuracy for better speed and efficiency.

Paper 6 introduced a multitask learning approach, combining fake news detection, classification, and risk scoring. This kind of setup is technically powerful because it allows the model to learn more generalizable features across tasks. However, it also increases model complexity and demands more high-quality labeled data.

Another thing I noticed is that most of the papers didn't go deep into explainability tools like SHAP or LIME, which are essential if we want to understand why the model made certain predictions. Only Paper 4 attempted to make the model more interpretable, but even that was limited to Chinese text.

So overall, while the models perform well in terms of accuracy, many of them lack transparency and are hard to implement efficiently in real-world environments.

Connections:

1. All six papers share a common objective: using artificial intelligence to detect and combat fake news, despite differences in language, datasets, and approaches.
2. Papers 1 and 2 mainly focus on general AI models, especially deep learning and large language models like BERT and RoBERTa, which are highlighted for their strong performance.
3. Paper 3 looks beyond just the content and explores user behavior, such as emotional tone and word repetition, which adds another useful dimension to fake news detection.
4. Paper 4 emphasizes explainability, offering insights into how models can be made more transparent and understandable to end users.
5. Papers 5 and 6 focus specifically on Arabic fake news, addressing a gap in research and introducing models (like AraBERT and MARBERTv2) that are adapted to the Arabic language.
6. All the studies aim to enhance accuracy in fake news detection, even though they use different methods — ranging from basic machine learning to multitask deep learning.
7. Some studies also point to the importance of adapting models for real-time use on platforms like social media, highlighting the need for flexibility and speed.

Conflict between finding:

1. While transformer-based models like BERT and MARBERTv2 are praised in several papers for their high performance, Paper 3 suggests that traditional ML models such as Random Forest can perform just as well when user behavior features are included.
2. Paper 6 introduces multitask learning (detection + classification + risk assessment), whereas other studies focus only on binary classification, favoring speed over detailed insights.
3. Some papers rely on pre-trained language models and embeddings, while others prefer hand-crafted features based on behavioral or linguistic cues.
4. Paper 5 promotes hybrid models (CNN + LSTM) for higher accuracy, but other research leans toward simpler models to avoid overfitting and complexity.
5. Evaluation methods vary certain studies prioritize accuracy and F1 score, while others focus more on explainability or how well the models apply in real-world scenarios.

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 with Chain-of-Thought and In-context Learning	Liu et al	2025	Explainability-Focused Study	Focused on making AI models more interpretable. Helps build user trust.	Used explainable AI methods (XAI) with simpler models.	Less accurate than deep models; some 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.

References:

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>

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

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>

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>

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>

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>