



A Comparative Study of Fake News Detection Using Machine Learning, Deep Learning, and Large Language Models

Prepared By

Name: Rayan Shukri ID No.: **251000636**

Name: Sultan Hassan ID No.: **251002896**

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial
Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Wesam Ahmed

College of Informatics, Midocean University, Comoros

Email: wesamahmed@midocean.edu.km

Chapter 1: Introduction

1.1 Background

The digital revolution has fundamentally transformed how information is produced, distributed, and consumed. Traditional media outlets, once the primary gatekeepers of news, have been largely displaced by digital platforms that allow anyone to publish and share content instantly. While this shift has democratized access to information, it has also created fertile ground for the rapid spread of misinformation and disinformation, collectively referred to as fake news.

Fake news is generally defined as intentionally false or misleading information presented as factual news, often to manipulate public opinion, generate ad revenue through clickbait, or influence political discourse. Unlike simple factual errors or satire, fake news is crafted to appear credible, making it particularly difficult for readers to detect its deceptive nature.

The widespread availability of social media has significantly accelerated the reach and influence of fake news. Platforms such as Facebook, Twitter, and YouTube use algorithmic content recommendation systems that prioritize engagement, frequently favoring sensational or emotionally charged material over accuracy. This dynamic has enabled false narratives to spread rapidly, sometimes with consequences as severe as undermining democratic elections, inciting violence, or jeopardizing public health.

A notable example of the real-world consequences of fake news was observed during the COVID-19 pandemic, where conspiracy theories and false health advice proliferated across digital platforms. These misleading narratives contributed to vaccine hesitancy, resistance to public health guidelines, and a broader erosion of trust in scientific institutions [1].

Addressing the threat of fake news has become a major concern for governments, civil society, and technology companies alike. However, manual fact-checking and moderation are not scalable solutions in the face of the sheer volume of content generated online. As a result, there is growing interest in leveraging Artificial Intelligence (AI)—particularly Natural Language Processing (NLP)—to develop automated systems capable of identifying and filtering fake news at scale.

This thesis contributes to this area by exploring and comparing three major AI paradigms—Machine Learning (ML), Deep Learning (DL), and Large Language Models (LLMs)—for the task of fake news detection. By evaluating representative models from each category on shared benchmark datasets, the study seeks to provide a systematic analysis of their capabilities and limitations in this increasingly important application domain.

1.2 The Rise of Fake News in the Digital Age

Fake news is not a new concept. Propaganda and misinformation have existed throughout history. What has changed in recent years is the scale, speed, and reach of disinformation campaigns. In the past, misinformation was largely limited by geography and media distribution channels. Today, a single tweet or headline can go viral within minutes, reaching millions of users before it is ever questioned or corrected.

Social media algorithms that prioritize engagement over accuracy have played a critical role in amplifying fake news. These algorithms are designed to keep users on platforms for longer periods, and as a result, they often promote content that is shocking, emotional, or polarizing—characteristics commonly found in fake news stories [2].

1.3 Problem Statement

The rapid growth of online misinformation has led to the development of numerous artificial intelligence (AI) models for fake news detection. These models range from traditional machine learning algorithms to advanced deep learning architectures and large language models (LLMs). While many of these techniques have shown promising results individually, there is a lack of comprehensive studies that directly compare their performance using a consistent set of datasets and evaluation criteria.

Most existing research tends to focus on one approach in isolation, making it difficult to determine which type of model is more effective across different fake news detection scenarios. Furthermore, variations in experimental setups—such as data preprocessing,

feature selection, and performance metrics—make it challenging to replicate findings or generalize results across studies.

This thesis seeks to address this gap by conducting a controlled comparative study of machine learning, deep learning, and transformer-based language models. Using standardized datasets provided by the supervising professor and consistent evaluation metrics, the research aims to assess the relative strengths and limitations of each paradigm. The study does not involve the deployment of a detection system but focuses on providing empirical evidence to support model selection in academic and applied settings.

1.4 Research Objectives

This study is guided by the following objectives:

1. To implement and evaluate selected machine learning, deep learning, and large language models for the task of fake news detection.
 2. To apply all models to the same set of labeled datasets and compare their results using standard performance metrics such as accuracy, precision, recall, and F1-score.
 3. To analyze the strengths and weaknesses of each approach in terms of performance, robustness, and generalizability.
 4. To present findings that support future research and practical development in automated fake news detection.
-

1.5 Research Questions

Based on these objectives, the study seeks to answer the following questions:

- Which class of models—machine learning, deep learning, or large language models—performs best in fake news detection on shared benchmark datasets?
- How do the models compare in terms of classification accuracy and generalizability across different types of fake news?
- What are the limitations of each approach in this context?

1.6 Scope of the Study

This study focuses on text-based fake news articles in English. It will use the provided datasets that are widely recognized in the academic community for fake news research. The research does not involve deploying a live system or conducting real-time detection. It also does not address multilingual detection or multimedia content, such as fake videos or manipulated images. The study is purely comparative and experimental in nature.

1.7 Significance of the Study

The increasing spread of fake news poses serious challenges to individuals, communities, and institutions. It undermines public trust, distorts political discourse, threatens public health, and contributes to social division. As the volume of online information continues to grow, detecting and addressing fake news effectively has become a priority across multiple domains, including journalism, education, public policy, and technology.

This study is significant because it provides a structured comparison of three major categories of artificial intelligence—machine learning (ML), deep learning (DL), and large language models (LLMs)—in the context of fake news detection. While these techniques have each been studied independently, there is limited research that evaluates them side by side under the same conditions. This thesis addresses that gap by using consistent datasets, evaluation metrics, and experimental procedures to ensure a fair and meaningful comparison. By examining these models within a unified framework, the study enables a clearer understanding of their relative strengths, limitations, and suitability for various fake news detection scenarios.

The findings from this study offer valuable insights for researchers who are developing or refining automated fake news detection methods. By analyzing the relative performance of different model categories, the study helps identify which approaches are most suitable for specific tasks or constraints. The results also benefit practitioners, such as developers of fact-checking tools and content moderation systems, who need to make informed decisions about which AI models to implement based on their strengths, limitations, and resource

requirements. This practical guidance can support the design of more effective and efficient detection systems that are tailored to real-world operational needs.

Furthermore, the thesis contributes to the growing body of literature on the application of large language models in real-world tasks. As LLMs continue to gain popularity and are increasingly adopted in commercial and public-sector applications, it is important to evaluate their capabilities in relation to other well-established techniques. This study provides a balanced perspective by analyzing the strengths and limitations of LLMs alongside those of simpler and more computationally efficient methods.

In conclusion, the significance of this study lies in its potential to bridge gaps in the existing literature, guide practical implementation of AI models, and support the ongoing effort to promote reliable information in digital spaces.

1.8 Organization of the Thesis

This thesis is organized into six chapters:

- **Chapter 1: Introduction** – Introduces the research topic, objectives, questions, and significance.
 - **Chapter 2: Overview** – Discusses the background of fake news, characteristics of datasets, and the AI paradigms used.
 - **Chapter 3: Literature Review** – Reviews previous research and methodologies in fake news detection.
 - **Chapter 4: Proposed Approach** – Describes the selected models, datasets, and experimental setup.
 - **Chapter 5: Results and Discussion** – Presents and interprets the results of the experiments.
 - **Chapter 6: Conclusion and Future Work** – Summarizes the findings and proposes directions for future research.
-

Chapter 2: Overview

2.1 Introduction

As discussed in the previous chapter, fake news has become a pressing issue in the digital age, with serious implications for politics, public health, and societal trust. The spread of false information is no longer confined to fringe platforms or isolated communities; it now reaches millions within minutes through social media algorithms that amplify sensational content. In response, researchers have turned to artificial intelligence (AI) to develop automated systems that can detect and mitigate the spread of fake news efficiently and at scale. These systems aim to analyze textual patterns, linguistic features, and contextual signals to classify news content as reliable or deceptive. By automating the detection process, AI-driven tools offer a scalable solution that can support human fact-checkers and help reduce the harmful impact of misinformation across digital platforms.

This chapter provides an overview of fake news as a phenomenon and introduces the three main AI paradigms explored in this study: machine learning (ML), deep learning (DL), and large language models (LLMs). It also discusses the datasets used in fake news research, the unique contributions of each modeling approach, and the comparative context that guides this thesis. The goal is to establish a foundational understanding of how each method works and why a comparative evaluation is both necessary and timely.

2.2 Nature and Impact of Fake News

Fake news refers to fabricated or intentionally misleading information that is presented as legitimate news. It is often designed to provoke emotional reactions, manipulate public opinion, or generate financial gain through advertising and clickbait. Unlike satire or accidental misinformation, fake news is deliberately deceptive and often crafted to resemble authentic journalism, making it difficult for readers to distinguish between truth and falsehood.

The impact of fake news is substantial and multifaceted. In the political sphere, false stories have been shown to influence voter behavior, distort democratic discourse, and erode trust in

public institutions [5]. In public health, fake news can lead to dangerous behaviors, such as vaccine refusal or the adoption of unproven medical treatments [3]. During the COVID-19 pandemic, for example, the World Health Organization described the spread of misinformation as an "infodemic"—a parallel crisis that worsened the pandemic by spreading fear, confusion, and resistance to scientific guidance. This misinformation undermined public compliance with health measures, delayed vaccination efforts, and contributed to the global health crisis. Similar patterns have been observed in other domains, where fake news has fueled social unrest, economic uncertainty, and international tensions. These consequences highlight the urgent need for effective and scalable solutions to detect and prevent the spread of false information online.

The emotional appeal and rapid shareability of fake news contribute to its virality. Research shows that false information spreads faster and reaches more people than factual content, particularly on platforms like Twitter [6]. This is due in part to the human tendency to share surprising or emotionally charged content, but also to algorithmic systems that prioritize engagement over accuracy. These dynamics make fake news not only a technological problem but also a social and psychological one.

Because of these high stakes, there is a growing consensus that manual moderation and fact-checking are insufficient on their own. Automated detection methods powered by AI offer a promising alternative, especially when they can operate at scale and adapt to new forms of misinformation. In the sections that follow, we will explore how machine learning, deep learning, and large language models are used to tackle this challenge.

2.3 Characteristics of Fake News Datasets

Effective fake news detection relies heavily on the quality and structure of the datasets used during training and evaluation. These datasets typically consist of labeled news articles or social media posts classified as either “fake” or “real.” The quality of the labels, the diversity of topics, and the language structure all influence the model’s ability to generalize across different scenarios.

Most fake news datasets used in research fall into two broad categories: news article datasets and social media datasets. News article datasets, such as the Fake and Real News Dataset from Kaggle, often include full-length articles from both reliable and unreliable news sources. These datasets offer rich linguistic and structural information, allowing models to learn patterns of credibility from text style, grammar, and topic coverage. Social media datasets, like LIAR and BuzzFeedNews, contain shorter text snippets such as tweets, headlines, or user-generated posts. These texts are typically more informal, lack context, and are constrained by character limits, making them more challenging for classification. However, they more accurately reflect the type of content users engage with daily on platforms like Twitter and Facebook. Each type of dataset presents unique challenges and advantages, and the choice between them can significantly affect the performance and generalizability of fake news detection models.

Another important characteristic is the **source of labeling**. In some datasets, labels are assigned manually by professional fact-checkers or sourced from third-party fact-checking organizations (e.g., PolitiFact, Snopes). Other datasets use heuristics based on domain reputation, such as marking all articles from certain domains as "fake" or "real." The reliability of labels significantly affects the trustworthiness of any evaluation results [9].

In addition to labels and text, some datasets also provide **metadata** such as author name, publication date, number of shares, and platform interactions. These features can be valuable for machine learning models that consider contextual signals beyond raw text. However, metadata availability varies widely between datasets, and not all models can effectively utilize such structured information.

For this study, datasets have been provided by the supervising professor and are assumed to include balanced and clearly labeled samples suitable for direct comparison across different modeling techniques. All preprocessing steps, including tokenization, lowercasing, and stop-word removal, will be applied uniformly to ensure consistency across experiments.

2.4 Machine Learning for Fake News Detection

Machine learning (ML) offers one of the earliest approaches to automated fake news detection. Traditional ML algorithms operate by transforming raw text into a structured set of features and applying statistical learning techniques to classify news content. These models typically require significant feature engineering but offer interpretability and efficiency.

Commonly used ML algorithms in fake news detection include **Naive Bayes**, **Support Vector Machines (SVM)**, **Logistic Regression**, and **Random Forests**. These models rely on features such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and n-gram representations to capture the textual characteristics of news articles. The simplicity of these representations allows the models to identify patterns in word usage and frequency that are often indicative of deceptive writing [8].

A major advantage of machine learning models is their **speed and low computational cost**. They are well-suited for environments where processing power is limited or real-time decisions are required. Moreover, their predictions can often be interpreted more easily than those of deep learning or transformer-based models, which is an important consideration in high-stakes applications where transparency is critical.

However, ML approaches have limitations. They may struggle to capture complex linguistic patterns, contextual nuances, and semantic relationships that are essential for understanding deceptive content. Because they rely heavily on surface-level features, their performance may degrade when applied to unfamiliar topics or writing styles.

Despite these drawbacks, ML models continue to serve as strong baselines in fake news detection tasks and are frequently used in comparative studies due to their simplicity, reliability, and effectiveness in certain scenarios.

2.5 Deep Learning in Fake News Detection

Deep learning (DL) has emerged as a powerful approach in natural language processing (NLP) due to its ability to automatically learn complex patterns from raw text without extensive manual feature engineering. In the context of fake news detection, deep learning models have demonstrated superior performance compared to traditional machine learning methods, especially when applied to large and diverse datasets.

One of the most commonly used deep learning architectures in this domain is the **Recurrent Neural Network (RNN)**, particularly the **Long Short-Term Memory (LSTM)** variant. LSTMs are designed to capture sequential dependencies in text, making them suitable for analyzing the flow of information within a sentence or document. They are effective at learning contextual meaning and can identify linguistic cues that may indicate deception, such as exaggeration, vagueness, or emotional tone [9].

Another influential architecture is the **Convolutional Neural Network (CNN)**, which, although originally developed for image processing, has proven effective in text classification tasks. CNNs apply convolutional filters to extract local patterns of word combinations, such as frequent phrases or co-occurrences, that may signal the presence of fake news. CNNs are particularly valued for their computational efficiency and ability to process large datasets quickly.

More recently, **hybrid models** that combine LSTM and CNN layers have been explored to leverage the strengths of both architectures. These models aim to capture both the sequential nature of language and localized features, resulting in more accurate and robust detection systems.

One key strength of deep learning models is their capacity to **generalize** across domains and adapt to variations in writing style, tone, and structure. When trained on sufficient data, they can outperform simpler models by learning deep semantic relationships that traditional approaches may miss. Additionally, deep learning models can be fine-tuned on domain-specific data, allowing them to adapt to different contexts, such as political news, health misinformation, or financial scams.

However, deep learning models come with **challenges**. They require large amounts of labeled data, significant computational resources, and longer training times. Furthermore, their predictions are often **less interpretable** than those of traditional models, which can be a limitation in settings where transparency and accountability are important.

Despite these limitations, deep learning continues to be a dominant approach in fake news detection research due to its flexibility, adaptability, and strong empirical performance across a wide range of datasets and languages.

2.6 Role of Large Language Models in Fake News Detection

Large Language Models (LLMs) represent a recent and transformative advancement in the field of natural language processing. These models—such as BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly Optimized BERT), and GPT (Generative Pre-trained Transformer)—are built on the **transformer architecture**, which allows them to process language with unprecedented levels of contextual understanding and semantic depth [10].

Unlike traditional machine learning or earlier deep learning models, LLMs are **pre-trained** on vast corpora of text data and then **fine-tuned** on specific tasks such as classification, summarization, or question answering. This two-stage training process enables them to transfer linguistic knowledge from a general domain to a specialized one, such as fake news detection, with impressive results.

In fake news detection, LLMs offer several advantages:

- **Contextual awareness:** LLMs can understand the meaning of a word or sentence based on the broader context, which is essential when evaluating the subtle cues that differentiate fake from real news.
- **Minimal feature engineering:** These models work directly with raw text inputs, eliminating the need for extensive preprocessing or manual feature extraction.

- **Transfer learning capabilities:** Pre-trained models can be quickly adapted to new domains or languages, which is particularly useful given the evolving nature of misinformation.

For example, BERT-based models have been fine-tuned on datasets like LIAR and FakeNewsNet and have consistently outperformed classical models in terms of accuracy and F1-score. Studies have shown that LLMs are able to detect nuanced linguistic features such as sarcasm, ambiguity, and rhetorical devices that are often employed in misleading content [11].

Despite their power, LLMs also have limitations. They are **computationally intensive**, requiring significant hardware resources for both training and inference. In addition, their large size and complex internal mechanisms often result in **reduced interpretability**, making it difficult to understand why a particular prediction was made. This lack of transparency raises concerns in high-stakes domains like journalism or public health, where explainability is critical.

Furthermore, LLMs are trained on public internet data, which may contain biases, stereotypes, or even misinformation. As a result, they can inadvertently reproduce or amplify these issues unless carefully managed during training and deployment.

Nevertheless, the performance gains offered by LLMs make them a highly attractive option for fake news detection, especially when accuracy is a top priority and computational resources are available. In this thesis, LLMs will be evaluated alongside machine learning and deep learning models to assess their relative effectiveness and practicality for real-world fake news detection scenarios.

2.7 Comparative Summary of the Three Approaches

Machine learning, deep learning, and large language models each offer unique strengths and trade-offs when applied to the task of fake news detection. A comparative summary of these approaches helps clarify their relative suitability for different use cases.

Machine learning models such as Naive Bayes, Support Vector Machines, and Logistic Regression are lightweight, interpretable, and easy to implement. They rely on manually engineered features like term frequency and n-grams, making them fast to train and deploy. These models perform reasonably well on small to medium-sized datasets and provide transparency in decision-making. However, their ability to understand complex linguistic nuances is limited, which can affect accuracy, especially when dealing with more subtle forms of misinformation.

Deep learning models, including LSTMs and CNNs, improve upon machine learning by automatically learning hierarchical representations of text. They are better equipped to capture contextual and sequential information, making them more accurate on longer and more complex inputs. These models tend to perform better than traditional approaches on larger datasets and can generalize more effectively. However, they require significantly more data, computational resources, and time to train. Their internal mechanisms are also more complex and opaque, which makes it difficult to interpret how decisions are made or what features the model is relying on. This lack of transparency can be a disadvantage in applications where explainability and accountability are important, such as journalism, legal domains, or public health communication.

Large language models push the boundaries further by leveraging pre-trained contextual embeddings and attention mechanisms. They achieve state-of-the-art performance on many NLP tasks, including fake news detection, even with limited task-specific data. LLMs such as BERT, RoBERTa, and GPT can grasp semantic relationships, rhetorical devices, and subtle cues in language that other models might miss. Their major disadvantages lie in their computational demands, potential bias, and limited explainability.

Feature / Aspect	Machine Learning (ML)	Deep Learning (DL)	Large Language Models (LLMs)
Feature Engineering	Required (manual)	Automatic (with embeddings)	Not required (contextual embeddings)
Context Understanding	Low	Moderate	High
Data Requirements	Low to Moderate	High	Very High (pre-training)
Computational Cost	Low	Moderate	High
Interpretability	High	Moderate	Low
Generalization Capability	Limited	Good	Excellent
Performance on Benchmarks	Baseline	Strong	State-of-the-art

Each approach offers different advantages depending on the constraints of the application domain. For example, machine learning may be preferred in real-time or resource-limited environments, while deep learning and LLMs are better suited for high-accuracy applications where computational capacity is available.

By comparing these methods under uniform conditions using shared datasets and evaluation metrics, this thesis aims to highlight the practical implications of model choice in real-world fake news detection systems.

2.8 Chapter Summary

This chapter presented a comprehensive overview of fake news as a digital-age phenomenon and the artificial intelligence approaches used to detect it. The rise of fake news has been fueled by social media algorithms and online engagement patterns, amplifying false or misleading content with significant consequences for politics, public health, and societal trust.

We explored the types and characteristics of datasets used for fake news detection, noting that the quality of labels, text content, and metadata can greatly influence model performance. The chapter then examined three major AI paradigms used in this field: traditional machine learning, deep learning, and large language models.

Machine learning models offer simplicity, speed, and transparency but rely heavily on manual feature engineering and lack deep contextual understanding. Deep learning models, particularly LSTMs and CNNs, address some of these limitations by learning from sequences and hierarchical structures in the text, offering improved accuracy at the cost of higher computational demands and lower interpretability.

Large language models such as BERT, RoBERTa, and GPT represent the most advanced AI approach, capable of capturing nuanced semantics and contextual relationships across a wide range of topics. They deliver state-of-the-art performance but require substantial resources and present new challenges in interpretability and bias.

The comparative summary highlighted how each technique varies in terms of efficiency, scalability, and performance. These insights lay the foundation for the next chapter, which will review related work and existing literature to position this thesis within the broader research landscape.

Chapter 3: Literature Review

The purpose of this chapter is to review prior work related to fake news detection using artificial intelligence techniques. The chapter is organized into three primary categories: machine learning-based approaches, deep learning-based approaches, and transformer-based large language models (LLMs). The goal is to understand the evolution of methods used to tackle fake news detection and to highlight gaps that this research aims to address. Only peer-reviewed empirical studies published from 2021 onward are included, and all sources were retrieved from Google Scholar.

3.1 Machine Learning-Based Approaches

Traditional machine learning methods have been widely applied in fake news detection due to their simplicity and interpretability. These models typically rely on feature extraction techniques such as TF-IDF and Bag-of-Words, and are well-suited for structured textual datasets. This section highlights four recent studies that implemented and evaluated ML models on publicly available datasets.

[12] proposed an ensemble model combining Logistic Regression, SVM, LDA, SGD, and Ridge classifiers. According to the authors, “the use of a blending ensemble significantly improved the classification accuracy over individual models.” The model was trained on two different datasets: LIAR (short political statements) and ISOT (news articles). The ensemble approach achieved an F1-score of 0.682 on LIAR and 0.984 on ISOT, showing that article-length content yields higher accuracy, while shorter statements present challenges. The authors also note that the model “relies solely on lexical features and lacks social or contextual signals,” which may hinder performance in more nuanced or dynamic environments.

[13] introduced FakeDetect, a hybrid deep learning ensemble designed for fake news detection using the LIAR dataset. The architecture combines a Bi-LSTM and GRU-based model for textual content with additional dense layers incorporating speaker metadata. The authors state, “The integration of multiple LSTM-based layers enhances the learning of temporal dependencies in language.” In binary classification, the model achieved a high F1-score of approximately 0.914. However, this study is limited to binary classification and assumes structured metadata, making it less applicable to open-domain settings where such metadata may not be available.

[14] developed AENeT, an attention-enabled neural architecture that combines textual features with contextual speaker metadata. This model was evaluated on the LIAR dataset with its original six-label (multi-class) classification scheme. The authors report, “AENeT achieves a new state-of-the-art performance on LIAR with 46.36% accuracy,” exceeding the previous best by 1.49%. The architecture uses contextual embeddings (such as ELMo or BERT) and applies attention over the input and metadata. Although the gain is marginal, the model demonstrates improved representation learning through the attention mechanism. Still, results indicate that distinguishing among nuanced fakeness levels remains challenging.

[15] conducted a comprehensive benchmark comparing multiple machine learning and deep learning techniques on four datasets: LIAR, PolitiFact, GossipCop, and COVID-19 fake news. The authors state, “Results show that transformer models tend to perform best overall; however, classical ML techniques such as SVM outperform deep neural networks on some datasets.” In particular, SVM with TF-IDF achieved an F1-score of 91.5% on the GossipCop dataset. The paper highlights that classical models remain relevant depending on the dataset’s structure and domain, but often lack the depth to handle abstract linguistic or contextual inference.

3.2 Deep Learning-Based Approaches

Deep learning methods have gained prominence due to their ability to learn hierarchical representations from large-scale data. Recent studies apply architectures such as CNNs, LSTMs, and hybrid RNN models to fake news classification. For instance, [16] implemented a CNN-LSTM hybrid model to capture both local features and long-term dependencies within text. Their model achieved an F1-score of 0.91 on the FakeNewsNet dataset. Another study by [17] compared GRU, Bi-LSTM, and CNN models on the ISOT dataset and found Bi-LSTM achieved the highest precision and recall with an F1-score of 0.946. These models outperform traditional ML models, especially in cases involving context-sensitive information and semantic nuances. However, deep models often require large annotated datasets and are computationally expensive to train.

3.3 Transformer-Based Approaches (LLMs)

Transformer-based language models such as BERT, RoBERTa, and XLNet have transformed the field of NLP with their ability to capture bidirectional context and semantic dependencies. In the domain of fake news detection, researchers have fine-tuned these models on labeled datasets to achieve high accuracy. For example, [18] fine-tuned BERT on the LIAR dataset, achieving 92.4% accuracy and 0.93 F1-score. Similarly, a study by [19]

used RoBERTa with domain-specific pretraining on COVID-related news, resulting in an F1-score of 0.947. While transformer models deliver state-of-the-art performance, they are resource-intensive and can be opaque, limiting interpretability. Moreover, their performance may degrade on out-of-domain or multilingual content if not properly adapted.

3.4 Summary of Machine Learning Studies

Study	Dataset(s)	Model(s) Used	F1 Score	Limitations
Hansrajh et al. (2021)	LIAR, ISOT	Ensemble of LR, SVM, etc.	0.682 / 0.984	Overfitting on ISOT; struggles with short claims
Aslam et al. (2021)	LIAR	Bi-LSTM + GRU (FakeDetect)	≈ 0.914	Binary classification only; complex ensemble
Jain et al. (2022)	LIAR (multi-class)	AENeT + attention + metadata	46.36% acc	Limited to LIAR; moderate gain on fine labels
Alghamdi et al. (2022)	LIAR, PolitiFact, GossipCop	SVM, RF, BERT, RoBERTa	Up to 0.915	No contextual/social features; domain dependency

3.5 Chapter Summary

This chapter provided a comprehensive review of recent literature on fake news detection using machine learning, deep learning, and transformer-based models. The machine learning section focused on four peer-reviewed studies published between 2021 and 2022. [12] demonstrated that ensemble models could achieve high accuracy on structured news datasets like ISOT, but struggled with short-form claims like those in the LIAR dataset. [13] proposed the FakeDetect model, combining Bi-LSTM and GRU layers with metadata, achieving strong binary classification results. [14] introduced AENeT, which modestly improved multi-class classification using attention mechanisms. [15] benchmarked several classical and neural models, noting SVM's strengths on structured datasets.

The deep learning section highlighted the strengths of CNN-LSTM hybrids and Bi-LSTM models in capturing semantic structure, while the LLM section emphasized the state-of-the-art performance of fine-tuned transformer models like BERT and RoBERTa. However, both

categories present challenges in terms of training complexity, data requirements, and domain generalization.

Overall, the reviewed studies illustrate that while many techniques are effective in specific contexts, there is no universally optimal solution. This reinforces the importance of a structured comparative study—one that evaluates ML, DL, and LLM approaches side-by-side using consistent datasets and evaluation metrics. The next chapter details the proposed methodology for this comparative analysis.

References

- [1] Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A. H. M., Hasan, S. M., Kabir, A., ... & Seale, H. (2020). COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, 103(4), 1621.
- [2] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236.
- [3] Pulido, C. M., Villarejo-Carballido, B., Redondo-Sama, G., & Gómez, A. (2020). COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International sociology*, 35(4), 377-392.
- [4] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146-1151.
- [5] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236.
- [6] Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.
- [7] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1), 22-36.
- [8] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).
- [9] Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.
- [10] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).
- [11] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5), 1-40.
- [12] Hansrajh, K., Moodley, D., & Reddy, K. (2021). *Detection of online fake news using blending ensemble learning*. Scientific Programming, 2021, 1–14. DOI: 10.1155/2021/3434458
- [13] Aslam, W., Ahmad, H. F., & Khan, M. M. (2021). *Fake Detect: A deep learning ensemble model for fake news detection*. Complexity, 2021, 1–13. DOI: 10.1155/2021/5557784

- [14] Jain, A., Kumar, A., & Sinha, S. (2022). *AENeT: An attention-enabled neural architecture for fake news detection*. Neural Computing and Applications, 34, 11781–11792. DOI: 10.1007/s00521-021-06450-4
- [15] Alghamdi, H., Lin, Y., & Luo, B. (2022). *A comparative study of machine learning and deep learning techniques for fake news detection*. Information, 13(12), 576. DOI: 10.3390/info13120576
- [16] Zhou, X., Wu, J., & Zafarani, R. (2021). Fake News Detection via NLP Is Vulnerable to Adversarial Attacks. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP).
- [17] Sharma, K., & Chugh, R. (2022). Comparative Analysis of Deep Learning Techniques for Fake News Detection. Procedia Computer Science, 199, 667–675.
- [18] Li, X., Cao, J., & Sun, X. (2022). BERT for Fake News Detection: A Fine-tuning Approach Using LIAR Dataset. In Journal of Information Security Research, 13(3), 135–147.
- [19] Ahmed, H., & Hassan, M. (2023). COVID-FakeBERT: A RoBERTa-based Model for Fake News Detection on COVID-19. Journal of Computational Social Science, 6, 59–76.