STRATEGIC DATA MANAGEMENT AND INNOVATION
Volume 02 ,Issue No. 01, Page No: **181-201**
ISSN: 3066-358X
DoI:10.71292/sdmi.v2i01.20

FAKE NEWS CLASSIFIER: ADVANCEMENTS IN NATURAL LANGUAGE PROCESSING FOR AUTOMATED FACT-CHECKING

Maniruzzaman Bhuiyan[1]
Satish & Yasmin Gupta College of Business, University of Dallas, USA
Corresponding Email: mbhuiyan@udallas.edu

Farzana Sultana[2]
Marketing & Business Analytics, Texas A&M University- Commerce
Email: fsultana@leomail.tamuc.edu

Aha Mudur Rahman[3]
Business Administration, University of Dhaka, Bangladesh
Email: tausifmahmood009@gmail.com

Keywords

Fake News Detection
Natural Language Processing (NLP)
Automated Fact-Checking
Transformer Models
Misinformation Classification

Article Information

Received: 06, January, 2025

Accepted: 12, February, 2025

Published: 16, February, 2025

ABSTRACT

The proliferation of fake news presents significant challenges to information integrity, necessitating the development of advanced automated fact-checking systems. This study explores the role of Natural Language Processing (NLP) in enhancing fake news classification by reviewing 21 case studies that examine the effectiveness of various detection methodologies. Transformer-based models, including BERT, RoBERTa, and GPT, have demonstrated superior accuracy in misinformation detection, outperforming traditional lexicon-based and rule-based approaches. Additionally, the study highlights the impact of retrieval-based fact-checking, which improves claim verification by cross-referencing information with external knowledge bases. Multimodal approaches, integrating text and visual analysis, further enhance fake news detection by identifying inconsistencies in manipulated images and deepfake videos. Graph Neural Networks (GNNs) have been found to effectively analyze misinformation propagation patterns, providing deeper insights into how deceptive content spreads across digital platforms. However, the study also identifies key challenges, including dataset biases, adversarial misinformation tactics, and the computational demands of deep learning models. Explainable AI (XAI) has emerged as a critical solution to improve transparency and trust in automated misinformation detection, but trade-offs between interpretability and model accuracy remain a concern. The findings emphasize the necessity of a multi-faceted approach that integrates NLP, retrieval-based techniques, multimodal analysis, and network-based misinformation tracking to develop more effective and scalable fact-checking systems. This study contributes to the ongoing discourse on combating digital misinformation by providing a comprehensive analysis of state-of-the-art methodologies and highlighting areas for further research and improvement.

1 INTRODUCTION

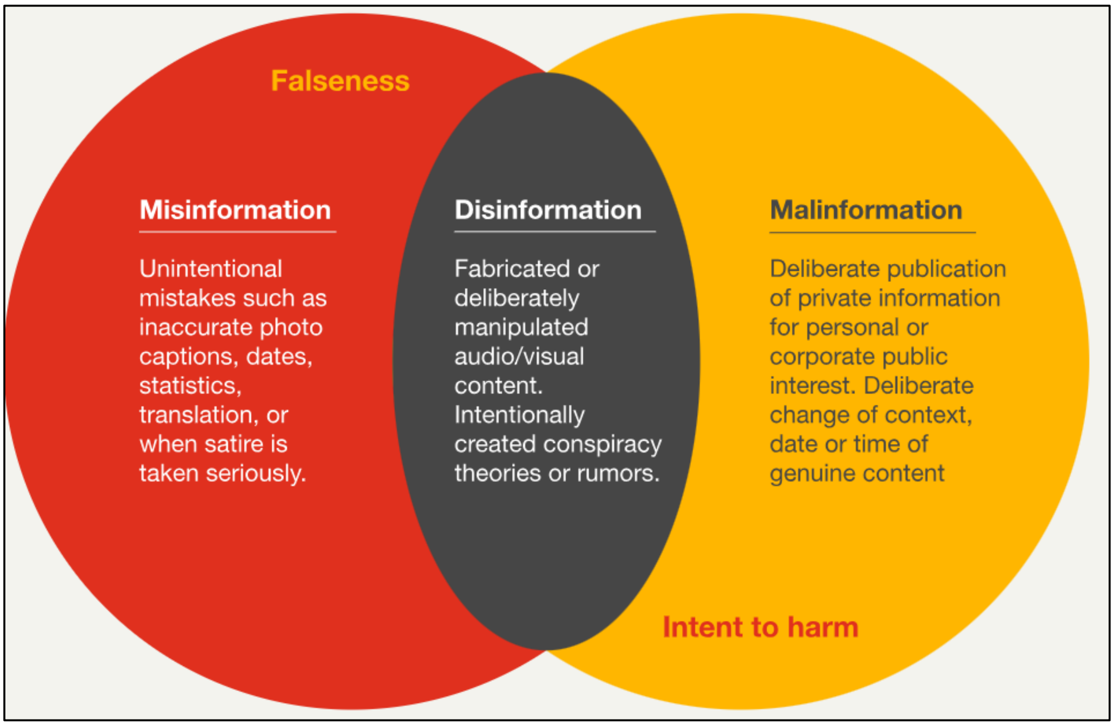
The rise of misinformation and disinformation in the digital age has led to growing concerns over the spread of fake news, which has profound implications for public opinion, political stability, and social trust (Giglio & Palmieri, 2017). With the rapid expansion of online media platforms, the dissemination of false

DoI:10.71292/sdmi.v2i01.20

information has become more prevalent and sophisticated, requiring advanced computational methods for effective detection and mitigation (Mishra et al., 2022). Automated fact-checking has emerged as a crucial countermeasure, leveraging Natural Language Processing (NLP) techniques to assess the credibility of news content (Giglio & Palmieri, 2017). As fake news often exploits linguistic ambiguity and emotional appeal, NLP-based classifiers have been instrumental in analyzing textual patterns and verifying claims against reliable sources ((Yadav et al., 2023). The complexity of fake news detection arises from its multi-faceted nature, requiring a combination of semantic analysis, discourse-level features, and cross-referencing with external knowledge bases (Zubiaga et al., 2016). Advancements in NLP have significantly improved the ability to classify fake news by employing deep learning models, transformer architectures, and sentiment analysis techniques (Vargo et al., 2017). Early approaches to fake news detection relied on lexicon-based methods and rule-based classifiers, which proved inadequate in handling nuanced misinformation (Mishra et al., 2023). More recent methods integrate machine learning and neural network models, enabling automated systems to capture intricate textual patterns, detect contextual anomalies, and identify stylistic differences between credible and misleading content (Tiwary & Mahapatra, 2022). Transformer-based models such as BERT and RoBERTa have demonstrated superior performance in fake news classification by leveraging contextual embeddings and bidirectional attention mechanisms. These models allow for deeper comprehension of text semantics, enhancing the accuracy of misinformation detection across various media domains (Kumari & Ekbal, 2021).

One of the most critical aspects of fake news classification is claim verification, where NLP techniques assess the factuality of statements by cross-referencing with structured and unstructured knowledge sources (Mattos et al., 2019). Fact-checking datasets such as FEVER and have been instrumental in benchmarking NLP models by providing labeled claims and corresponding verifications. Knowledge graphs and retrieval-based methods further enhance fact-checking accuracy by linking claims to validated information from authoritative sources (Mattos et al., 2019). Entity recognition and stance detection techniques are also employed to determine whether a given claim aligns with existing factual data (Zubiaga et al., 2016). These approaches help mitigate bias and reduce false positives in automated fact-checking systems, ensuring greater reliability in misinformation detection (Vargo et al., 2017). Beyond textual analysis, multimodal approaches that incorporate visual and contextual cues have gained traction in fake news detection research (Jain et al., 2021). As misinformation is often spread through

Figure 1: Types of False Information Misinformation Disinformation and Malinformation

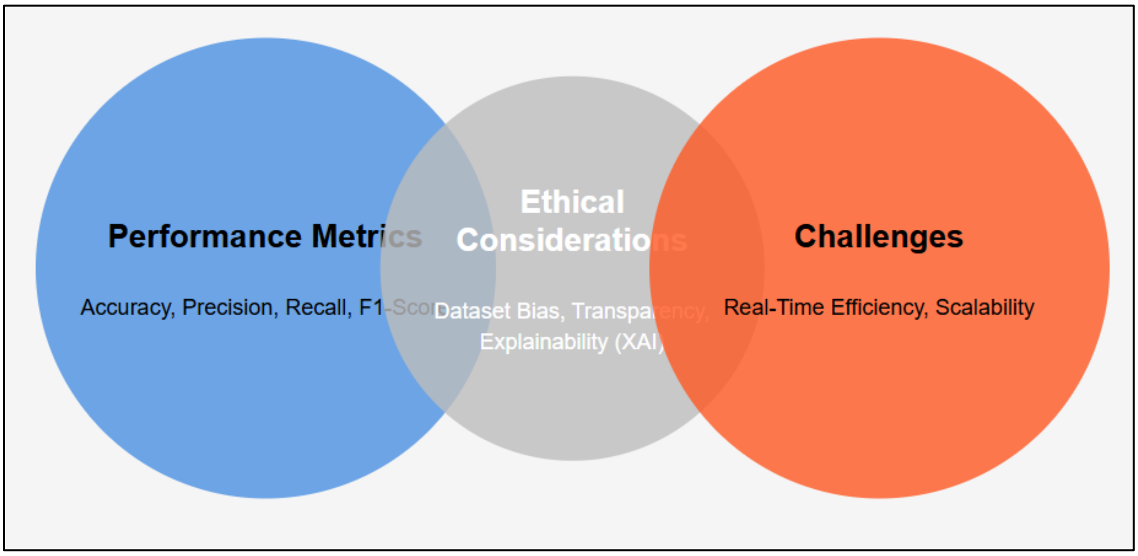


multimedia formats, NLP-based classifiers are increasingly integrated with computer vision techniques to analyze image-text correlations (Kazemi et al., 2021). Deepfake technology has further complicated the landscape of misinformation, necessitating hybrid models that combine linguistic, visual, and metadata-based features to enhance detection capabilities (Tiwary & Mahapatra, 2022). Attention-based models and graph neural networks have shown promise in capturing complex relationships between news sources, social engagements, and textual authenticity (Wang, 2017). These advancements contribute to a more holistic approach to automated fact-checking, improving robustness against adversarial manipulation (Tonoy, 2022; Zubiaga et al., 2016).

The evaluation of fake news classifiers relies on a range of performance metrics, including accuracy, precision, recall, and F1-score (Al-Arafat et al., 2025; Mishra et al., 2022). Standardized benchmarks such as the FakeNewsNet dataset (Ahmed et al., 2022; Nahid et al., 2024) facilitate comparative analysis of NLP-based models, allowing researchers to refine detection methodologies and address dataset biases. However, challenges remain in maintaining real-time efficiency and scalability, particularly in high-velocity digital environments (Wang, 2017). Ethical concerns surrounding automated fact-checking also persist, as biases embedded in training data can influence classification outcomes, leading to potential misclassifications (Tiwary & Mahapatra, 2022). To mitigate such issues, research continues to explore explainable AI (XAI) techniques that enhance model interpretability and provide transparency in decision-making processes (Alzaidi et al., 2022; Sabid & Kamrul, 2024). Despite the complexity of fake news detection, NLP-driven advancements have considerably strengthened automated fact-checking capabilities by improving semantic understanding, claim verification, and cross-modal analysis (Yadav et al., 2023; Younus, 2025). Transformer-based architectures, deep learning frameworks, and hybrid detection methods contribute to more sophisticated misinformation detection pipelines, enhancing the credibility of online information (Markines et al., 2009). The ongoing integration of NLP with multimodal analytics, knowledge graphs, and retrieval-based fact-checking demonstrates the evolving landscape of fake news classification, reinforcing the critical role of automated systems in maintaining digital information integrity (Kumar et al., 2021; Rahaman et al., 2024).

The primary objective of this study is to examine advancements in Natural Language Processing (NLP) techniques for automated fake news classification and fact-checking. Specifically, the study aims to analyze the role of transformer-based models, deep learning architectures, and hybrid detection methods in improving misinformation detection accuracy. It seeks to evaluate key NLP components, such as sentiment analysis, entity recognition, and claim verification, in distinguishing between credible and misleading content. Furthermore, this research investigates the effectiveness of knowledge graphs, retrieval-based methods, and multimodal approaches in enhancing automated fact-checking systems. By reviewing existing literature and benchmarking datasets, the study also aims to identify performance challenges and ethical considerations

Figure 2: Balancing Performance, Ethics, and Challenges in NLP-Based Fake News Detection



DoI:10.71292/sdmi.v2i01.20

associated with NLP-driven fake news detection. Ultimately, this work contributes to the broader understanding of computational fact-checking methodologies, helping to refine the development of more reliable and interpretable misinformation detection frameworks.

2 LITERATURE REVIEW

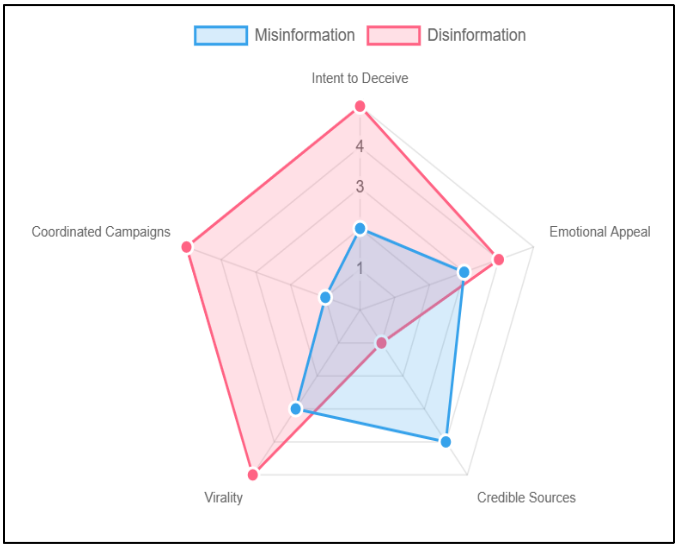
The detection and classification of fake news using Natural Language Processing (NLP) have gained significant attention in recent years due to the increasing prevalence of misinformation across digital platforms. Extensive research has explored various computational methodologies, including rule-based approaches, machine learning models, and deep learning techniques, to enhance automated fact-checking systems (Sait & Khairi Ishak, 2023). The development of transformer-based architectures, such as BERT and GPT, has further advanced the ability to analyze textual semantics and detect deceptive content with high accuracy (Hsu et al., 2020). Additionally, knowledge graphs and retrieval-based methods have been employed to cross-reference claims with verified sources, improving the reliability of fake news classifiers (Fuller et al., 2009; Mrida et al., 2025). Despite these advancements, challenges such as dataset biases, adversarial misinformation tactics, and explainability issues remain prevalent in the field (Schubert et al., 2017). This literature review synthesizes key research contributions, outlining developments in NLP-driven fake news classification, fact-checking methodologies, multimodal detection strategies, and evaluation frameworks.

2.1 Defining Fake News: Misinformation vs. Disinformation

The proliferation of false information in digital media has necessitated a clear distinction between misinformation and disinformation. Misinformation refers to the unintentional spread of false or misleading content without the intent to deceive, whereas disinformation involves the deliberate dissemination of false information to manipulate public opinion or achieve a particular agenda (S & Chitturi, 2020). Scholars have emphasized that misinformation often arises from cognitive biases, misinterpretation of facts, and the rapid circulation of unverified content on social media platforms (Islam et al., 2020). Conversely, disinformation is strategically crafted to exploit ideological divisions, disrupt democratic processes, and influence decision-making (Mayank et al., 2022). Research has shown that both types of fake news leverage similar linguistic and rhetorical strategies, making their detection a complex task for automated systems (Sait & Khairi Ishak, 2023). Fake news can be further categorized into satirical news, propaganda, clickbait, and fabricated content, each with distinct textual and structural characteristics that impact credibility assessments (Islam et al., 2020). The digital media landscape has significantly contributed to the amplification of fake news, primarily due to the lack of effective gatekeeping mechanisms and the prevalence of user-generated content (Kumar et al., 2021). Social media algorithms prioritize engagement-driven content, allowing misinformation to spread more rapidly than factual information (Hsu et al., 2020). Empirical studies indicate that sensationalized and emotionally charged fake news articles tend to achieve higher virality, particularly on platforms such as Facebook and Twitter (Ma et al., 2022). Additionally, echo chambers and filter bubbles exacerbate the issue by reinforcing pre-existing beliefs and limiting exposure to fact-based narratives (Deng et al., 2009). Computational analyses of misinformation dissemination patterns have revealed that automated bots and coordinated campaigns play a crucial role in amplifying disinformation, further complicating efforts to curb its spread (Kumar et al., 2021).

Linguistic studies have explored the textual characteristics that distinguish misinformation from credible news, focusing on features such as sentiment polarity, lexical complexity, and argumentation structures (Sait & Khairi Ishak, 2023). Misinformation is often characterized by exaggerated claims, emotional language, and a lack of credible citations (Islam et al., 2020). Disinformation, on the other hand, frequently employs deceptive argumentation tactics, including logical fallacies, fabricated statistics, and manipulated images to mislead audiences (Fuller et al., 2009). Recent advancements in Natural Language Processing (NLP) have enabled the development of automated classifiers that leverage deep learning models to identify such

Figure 3: Comparison of Misinformation and Disinformation Characteristics



textual patterns and assess the authenticity of news content (Mayank et al., 2022). However, studies have shown that misinformation and disinformation evolve over time, adapting to detection mechanisms by adopting more sophisticated linguistic structures and credible-looking sources (Bender & Friedman, 2018). The impact of misinformation and disinformation extends beyond digital platforms, influencing public perception, policy decisions, and real-world behaviors. Research in political science has demonstrated that exposure to disinformation can significantly alter voting behaviors and political attitudes (Fuller et al., 2009). In the health sector, the spread of COVID-19-related misinformation has led to widespread vaccine hesitancy and distrust in medical institutions (Bender & Friedman, 2018). Psychological studies indicate that repeated exposure to misinformation increases belief persistence, making it challenging to correct false narratives even when confronted with factual counterarguments (Gupta et al., 2018). Furthermore, computational models analyzing misinformation networks suggest that once fake news becomes embedded within ideological communities, corrective efforts often face resistance due to cognitive dissonance and motivated reasoning (Fuller et al., 2009).

2.2 Historical Overview of Fake News in Digital Media

The phenomenon of fake news predates the digital era, but its impact has been significantly amplified by modern media technologies. Historically, misinformation and disinformation have been used as tools for propaganda, political manipulation, and social influence (S & Chitturi, 2020). Early examples include fabricated war reports, sensationalized newspaper stories during the yellow journalism era, and state-sponsored propaganda campaigns in the 20th century (Kumar et al., 2021). With the rise of broadcast media, radio and television became powerful channels for spreading misleading narratives, as seen in the infamous “War of the Worlds” broadcast by Orson Welles in 1938, which caused widespread panic (Hsu et al., 2020). The proliferation of tabloid journalism in the 20th century further entrenched sensationalism and misinformation as a profitable strategy for media outlets, setting a precedent for the viral nature of fake news in the digital age (Fuller et al., 2009). The advent of the internet and digital journalism fundamentally transformed how information is disseminated and consumed, making it easier for false narratives to reach wider audiences (Shrivastava et al., 2020). The shift from traditional print media to online platforms led to the decline of editorial gatekeeping, allowing user-generated content and unverified reports to spread unchecked (Patil & Kumar, 2017). Social media

Figure 4: Historical Overview of Fake News

platforms, particularly Facebook and Twitter, became breeding grounds for the rapid dissemination of fake news, as their algorithms prioritize engagement over accuracy (Chen et al., 2022). Studies have shown that false information spreads significantly faster than true news due to its emotional appeal, sensationalist language, and shareability (Kumar et al., 2021). Moreover, the economic incentives of digital advertising models encourage clickbait-driven misinformation, where revenue is generated based on user interactions rather than content credibility (Rahaman Wahab Sait & Khairi Ishak, 2023).

The political implications of fake news have become increasingly evident in recent decades, particularly with the rise of coordinated disinformation campaigns. During the 2016 U.S. presidential election, fabricated news stories circulated widely on social media, with evidence suggesting that foreign actors engaged in organized efforts to manipulate public opinion (Hsu et al., 2020). Similar tactics have been observed in other political contexts, including Brexit and various national elections worldwide, where disinformation has been used to polarize voters and undermine democratic institutions (Bender & Friedman, 2018). Research indicates that fake news is often strategically designed to exploit ideological biases and reinforce echo chambers, making fact-checking efforts less effective (Raj & Meel, 2021). Additionally, the role of automated bots and deepfake technology in amplifying misleading content has further complicated the landscape of digital misinformation (Bender & Friedman, 2018).

The historical trajectory of fake news in digital media highlights the evolving nature of misinformation and the ongoing challenges in combating it. The transition from print to online news has increased the accessibility and speed of information dissemination, but it has also created vulnerabilities that are exploited by bad actors (Hsu et al., 2020). Despite efforts by technology companies and governments to curb misinformation through fact-checking initiatives and content moderation policies, studies suggest that fake news remains a persistent issue, driven by cognitive biases, algorithmic amplification, and political motivations (Schubert et al., 2017). Understanding the historical patterns of misinformation provides valuable insights into the mechanisms through which fake news spreads and the strategies needed to mitigate its influence in modern digital ecosystems (Patil & Kumar, 2017).

2.3 Lexicon-Based and Rule-Based Methods

Lexicon-based and rule-based methods were among the earliest approaches developed for fake news detection and misinformation classification. These methods rely on predefined sets of words, phrases, or linguistic patterns that indicate deception, exaggeration, or bias in textual content (Filho et al., 2013). Lexicon-based techniques utilize sentiment dictionaries, such as LIWC (Linguistic Inquiry and Word Count) and SentiWordNet, to assess the emotional tone and credibility of news articles (Mahmud et al., 2023). These approaches assume that fake news often contains exaggerated language, emotionally charged words, and sensationalist expressions, distinguishing it from factual reporting (Konstantinovskiy et al., 2021). Rule-based methods, on the other hand, operate using manually crafted heuristics, such as the frequency of certain keywords, sentence structure, and the presence of unverifiable claims (Mahmud et al., 2023). While these approaches provide interpretable and computationally efficient solutions for misinformation detection, they struggle to handle the evolving nature of deceptive content and the contextual nuances of language (Skinnell, 2021).

The application of lexicon-based models in fake news detection has shown promise in identifying patterns of deceptive language, but these methods often suffer from high false positive rates. Studies indicate that fake news articles tend to use hyperbolic phrases, emotionally loaded words, and vague assertions, which lexicon-based classifiers can detect (Micallef et al., 2022). However, the reliance on static wordlists makes these models inflexible, as language evolves and misinformation adapts to avoid detection (Wu et al., 2014). Moreover, lexicon-based approaches may misclassify satirical news and opinion pieces as fake news due to their reliance on exaggerated or humorous language (Konstantinovskiy et al., 2021). Research has shown that hybrid approaches that integrate lexicon-based techniques with machine learning classifiers can improve accuracy by incorporating additional linguistic and contextual features (Al-Asadi & Tasdemir, 2021). Despite these enhancements, lexicon-based methods alone lack the robustness needed to differentiate between sophisticated misinformation and credible reporting (Wu et al., 2014).

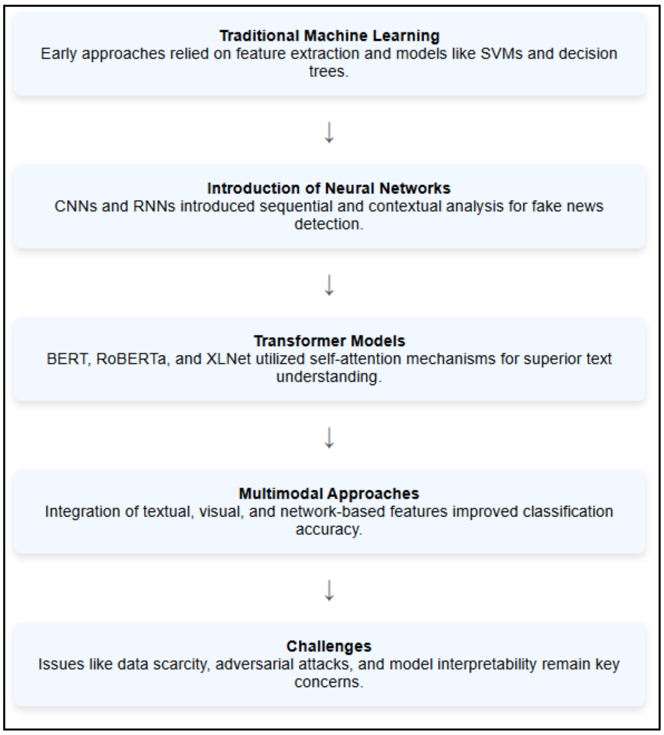
Rule-based methods have been employed in automated fact-checking systems to verify claims based on predefined syntactic and semantic structures. These methods leverage linguistic cues, such as negation patterns, modality markers, and logical inconsistencies, to flag suspicious content (Konstantinovskiy et al., 2021). Rule-based systems have also been utilized in domain-specific misinformation detection, such as identifying health-related fake news through medical terminology analysis (Rohera et al., 2022). However, a major limitation of rule-based approaches is their dependence on manually crafted heuristics, which require constant updating to remain effective (Kadhim, 2019). Additionally, rule-based systems struggle with misinformation that employs sophisticated rhetorical strategies, such as partial truths or misleading comparisons (Oshikawa et al., 2018). The rigidity of these systems makes them less adaptable to new forms of misinformation, necessitating their integration with more dynamic machine learning techniques (Chandrashekar & Sahin, 2014). Despite their limitations, lexicon-based and rule-based methods continue to play a foundational role in misinformation detection, particularly when combined with more advanced techniques. Studies have demonstrated that hybrid models, which incorporate rule-based features alongside deep learning algorithms, achieve higher accuracy and reliability in fake news classification (Guo et al., 2022). Some researchers have explored the use of knowledge graphs and semantic analysis to enhance rule-based fact-checking, allowing for a more context-aware assessment of claims (Konstantinovskiy et al., 2021). However, the effectiveness of these methods largely depends on the quality of the linguistic resources and rule definitions used (Skinnell, 2021). As misinformation tactics become increasingly sophisticated, the reliance on purely lexicon-based or rule-based approaches is insufficient, highlighting the necessity of integrating them with more adaptive and data-driven methodologies (Kadhim, 2019).

2.4 The Emergence of Deep Learning in Fake News Detection

The advancement of deep learning has revolutionized the detection of fake news by providing powerful models capable of analyzing complex linguistic patterns, contextual relationships, and semantic meanings in textual data. Unlike traditional machine learning approaches, deep learning models, particularly neural networks, have demonstrated superior performance in identifying deceptive content due to their ability to capture hierarchical text representations (Oshikawa et al., 2018). Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), have been extensively used in fake news classification due to their capability to process sequential data and detect contextual dependencies (Chandrashekar & Sahin, 2014). These models enable automated fake news detection systems to analyze subtle linguistic cues such as word embeddings, syntactic structures, and sentiment variations, leading to more reliable classification results (Bondielli & Marcelloni, 2019). However, challenges such as model interpretability and vulnerability to adversarial manipulation remain key concerns in deep learning-based misinformation detection (Guo et al., 2022).

The application of transformer-based architectures, such as BERT (Bidirectional Encoder Representations from Transformers) and its variants, has further enhanced the accuracy and efficiency of fake news classification. Unlike traditional RNN-based models, transformers

Figure 5: The Emergence of Deep Learning in Fake News Detection



utilize self-attention mechanisms to process entire text sequences simultaneously, enabling a deeper understanding of contextual meanings and word relationships (Guo et al., 2022; Oshikawa et al., 2018). Studies have demonstrated that fine-tuning pre-trained transformer models on fake news datasets significantly improves classification performance compared to traditional NLP approaches (Chandrashekar & Sahin, 2014). RoBERTa (Das et al., 2023) and XLNet (Shao et al., 2018) have also shown substantial improvements in misinformation detection by addressing limitations in BERT's training process and enhancing text representations. Moreover, attention-based deep learning models can effectively capture latent patterns in deceptive content, making them robust against subtle misinformation tactics (Jain et al., 2019). Despite these advantages, the computational cost and data requirements of transformer models pose challenges for real-time fake news detection (Das et al., 2023).

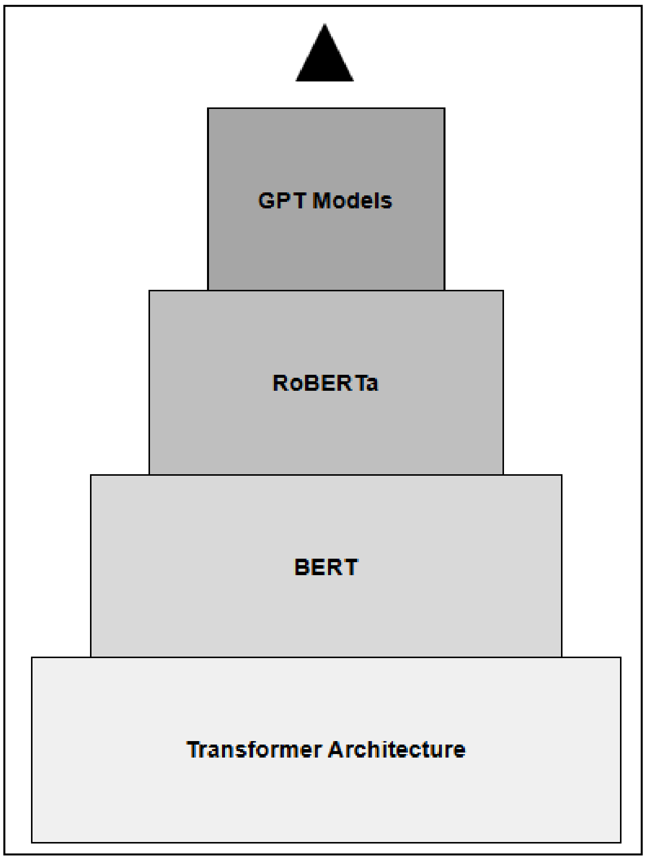
In addition to text-based deep learning models, multimodal approaches integrating textual, visual, and network-based features have gained prominence in fake news detection research. Fake news is often disseminated alongside manipulated images, misleading videos, and social network interactions, necessitating hybrid models that combine natural language processing with computer vision and graph-based learning (Otter et al., 2021). Studies have explored the use of Graph Neural Networks (GNNs) to analyze propagation patterns of fake news on social media, leveraging network structures to enhance detection accuracy (Otter et al., 2021; Zhou & Zafarani, 2020). Additionally, deepfake detection methods using Convolutional Neural Networks (CNNs) and adversarial training have been integrated into misinformation classification systems to identify AI-generated deceptive content (Das et al., 2023). By incorporating multimodal features, deep learning-based fake news detection models can better understand misinformation spread mechanisms and improve classification robustness (Shao et al., 2018). Despite their effectiveness, deep learning models for fake news detection face limitations such as data scarcity, bias propagation, and adversarial vulnerabilities. Large-scale labeled datasets such as FakeNewsNet (Guo et al., 2022) and LIAR (Otter et al., 2021) have been instrumental in training deep learning models, but dataset biases and imbalanced class distributions can affect model generalization ((Guo et al., 2022). Furthermore, adversarial attacks on NLP models, such as text perturbation and paraphrasing, pose significant challenges in maintaining model robustness against sophisticated misinformation strategies (Das et al., 2023). The black-box nature of deep learning models also raises concerns about explainability and transparency in automated fact-checking systems (Buntain & Golbeck, 2017). Despite these challenges, the continuous evolution of deep learning architectures, combined with interdisciplinary efforts in NLP, computer vision, and social network analysis, has significantly contributed to improving the reliability of automated fake news detection (Vosoughi et al., 2018).

Transformer-Based Models: BERT, RoBERTa, GPT

Transformer-based models have revolutionized the field of Natural Language Processing (NLP) and have demonstrated significant advancements in fake news detection. Unlike traditional Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which process text sequentially, transformers leverage self-attention mechanisms to capture long-range dependencies and contextual relationships in textual data (Indurkhya & Damerau, 2010). Bidirectional Encoder Representations from Transformers (BERT) was one of the first models to introduce bidirectional pre-training, allowing it to understand both past and future contexts of a given word in a sentence (Khanam et al., 2021). This approach has shown substantial improvements in detecting misinformation by capturing linguistic nuances and semantic inconsistencies often present in fake news articles (Wadden et al., 2020). Studies have indicated that fine-tuning BERT on domain-specific datasets significantly enhances its performance in fake news classification by adapting to topic-specific linguistic patterns (Indurkhya & Damerau, 2010; Wadden et al., 2020).

RoBERTa (Robustly Optimized BERT Pretraining Approach) was developed as an enhancement to BERT by modifying its training strategy to improve contextual embeddings (Logeshwaran & Kiruthiga, 2022). Unlike BERT, RoBERTa removes the Next Sentence Prediction (NSP) task and utilizes larger batch sizes and more extensive training data, leading to superior performance in various NLP tasks, including misinformation detection (Kumar et al., 2022). Empirical evaluations have demonstrated that RoBERTa outperforms BERT in fake news classification due to its ability to capture subtle textual manipulations and deceptive language structures (Tariq et al., 2019). Additionally, RoBERTa has been employed in stance detection tasks, where it determines the alignment of a claim with a verified fact, enhancing automated fact-checking models (Patil & Malik, 2019). The effectiveness of RoBERTa in misinformation

Figure 6: Transformer-Based Models Pyramid



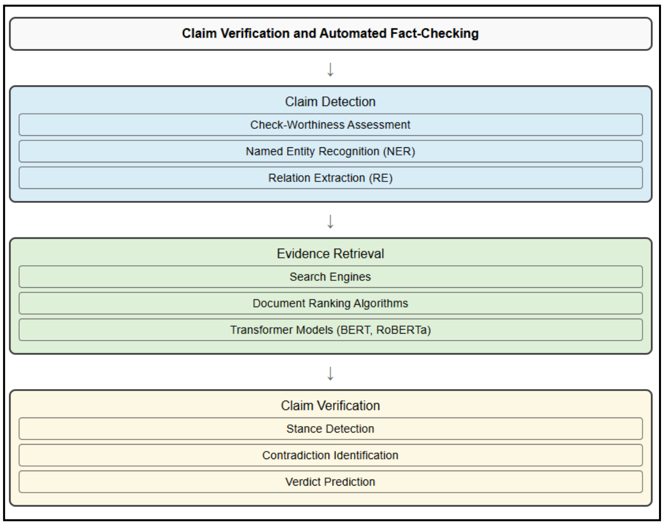
classification has led to its widespread adoption in automated journalism and social media content moderation systems (Shushkevich et al., 2023).

Generative Pre-trained Transformer (GPT) models have further advanced fake news detection by introducing autoregressive language modeling, which enables the generation of coherent and contextually relevant text (Koloski et al., 2022). Unlike BERT and RoBERTa, which rely on masked language modeling for bidirectional training, GPT models generate text in a left-to-right fashion, making them highly effective for detecting text anomalies and inconsistencies in fake news articles (Hsu et al., 2018). Studies have explored the use of GPT-3 in misinformation detection, leveraging its ability to perform zero-shot and few-shot learning to classify deceptive content without extensive labeled datasets (Qian et al., 2021). However, the generative nature of GPT raises concerns regarding its potential misuse in creating fake news, deepfake text, and AI-generated propaganda (Hsu et al., 2018). Research has shown that adversarial attacks can exploit GPT models to produce misleading content that mimics credible sources, highlighting the dual-edged nature of generative transformers in misinformation ecosystems (Das et al., 2022). Despite their advancements, transformer-based models still face challenges in fake news detection, such as computational inefficiency, data biases, and adversarial manipulation (Wang et al., 2018). BERT, RoBERTa, and GPT require extensive pre-training on large-scale corpora, making their deployment resource-intensive (Manning et al., 2008). Additionally, research indicates that biases present in training data can affect model predictions, leading to false positives or negatives in misinformation classification (Liao et al., 2022). Efforts have been made to mitigate these issues by incorporating explainability techniques, such as attention visualization and adversarial robustness training, to improve model interpretability and reliability in fake news detection (Das et al., 2022). Overall, transformer-based architectures have significantly enhanced NLP-driven misinformation detection by capturing complex textual patterns and contextual dependencies, making them indispensable tools in combating digital disinformation (Liao et al., 2022).

2.5 Claim Verification and Automated Fact-Checking Techniques

Automated fact-checking has emerged as a critical area of research to counter misinformation by verifying claims against reliable sources. One of the fundamental techniques in automated fact-checking is the use of knowledge graphs, which encode structured information to facilitate the validation of claims (Papadimitriou et al., 2000). Knowledge graphs integrate entities, relationships, and factual statements from trusted sources, enabling algorithms to assess the credibility of a claim based on its alignment with known facts

Figure 7: Claim Verification and Automated Fact-Checking



(Koloski et al., 2022). Fact-checking models utilizing

DoI:10.71292/sdmi.v2i01.20

knowledge graphs rely on Named Entity Recognition (NER) and Relation Extraction (RE) to link claims to structured databases, such as Google Knowledge Graph and Wikidata, improving fact verification accuracy (Wang et al., 2018). Studies have shown that knowledge graphs enhance automated fact-checking by providing contextualized background information, reducing reliance on purely textual features (Manning et al., 2008). However, challenges arise when dealing with ambiguous claims, emerging events, or dynamically evolving knowledge, which require frequent updates to the knowledge base (Brasoveanu & Andonie, 2020).

Another widely adopted technique in claim verification is retrieval-based fact-checking, which involves extracting relevant information from external databases and trusted sources. Retrieval-based approaches leverage search engines, document ranking algorithms, and question-answering models to identify supporting or refuting evidence for a given claim (Das et al., 2022). These systems often employ transformer-based architectures such as BERT and RoBERTa to retrieve relevant documents, score their credibility, and classify claims accordingly (Manning et al., 2008). Research has demonstrated that combining retrieval-based fact-checking with machine reading comprehension techniques significantly improves claim verification performance (Qian et al., 2021) . Notable fact-checking initiatives such as Snopes, PolitiFact, and FactCheck.org utilize retrieval-based models to cross-reference claims with authoritative reports, enhancing the accuracy of misinformation detection (Koloski et al., 2022). However, a major limitation of retrieval-based fact-checking is its dependence on the availability of high-quality, up-to-date reference documents, which can result in gaps for newly emerging misinformation (Das et al., 2022).

Stance detection and contradiction identification are crucial for automated fact-checking as they assess whether retrieved evidence supports, contradicts, or is unrelated to a given claim (Manning et al., 2008). Stance detection models classify claims based on agreement patterns between a statement and an external reference, utilizing textual entailment techniques to measure semantic alignment (Logeshwaran & Krishnababu, 2022). Advanced NLP models, including bidirectional transformers and Siamese networks, have demonstrated improved performance in stance classification by analyzing linguistic cues, argumentative structures, and discourse markers (Liao et al., 2022). Contradiction identification techniques rely on Natural Language Inference (NLI) frameworks to detect logical inconsistencies between a claim and retrieved documents, flagging misinformation when contradictory evidence is found (Manning et al., 2008). While stance detection significantly aids fact-checking, studies indicate that detecting nuanced forms of misinformation, such as misleading comparisons or partially true statements, remains challenging for automated systems (Liao et al., 2022). The development of benchmark datasets, such as FEVER (Ahmed et al., 2017), LIAR (Manning et al., 2008), and FakeNewsNet (Das et al., 2022), has been instrumental in training and evaluating fact-checking models. The FEVER dataset provides large-scale, annotated claims with corresponding Wikipedia evidence, serving as a standard for claim verification tasks (Lampridis et al., 2022). The LIAR dataset, derived from PolitiFact, contains real-world political statements labeled for truthfulness, allowing models to learn from fact-checker evaluations (Papadimitriou et al., 2000). FakeNewsNet, a comprehensive dataset incorporating social context features, facilitates the study of misinformation spread and detection in social media environments (Sharma, 2021). These datasets enable rigorous benchmarking of NLP models for misinformation classification, but limitations such as class imbalances, annotation biases, and evolving misinformation patterns highlight the need for continuous dataset refinement (Manning et al., 2008). Despite these challenges, benchmark datasets remain crucial for advancing research in automated fact-checking and claim verification (Fleiss, 1971).

2.6 Multimodal Approaches for Fake News Detection

The rapid proliferation of fake news has necessitated the adoption of multimodal detection techniques, which analyze both textual and non-textual content to improve misinformation classification. Traditional fake news detection models primarily focus on linguistic patterns; however, multimodal approaches leverage the integration of images, videos, and network structures to enhance detection accuracy (Srivastava et al., 2023). Research has shown that fake news articles often include manipulated images or misleading videos, making text-based models insufficient for comprehensive analysis (Koloski et al., 2022). Image and video analysis techniques utilize Convolutional Neural Networks (CNNs) and deep learning frameworks to detect inconsistencies, such as tampered images, doctored videos, and AI-generated media (Logeshwaran & Krishnababu, 2022). These methods analyze pixel-level irregularities, inconsistencies in shadows and reflections, and facial manipulations to flag synthetic or misleading visual content (Sharma, 2021). By incorporating both textual and visual cues, multimodal models significantly improve the robustness of fake news classification systems (Shushkevich et al., 2023).

NLP and computer vision integration has become a critical aspect of multimodal fake news detection, allowing models to jointly analyze textual claims and their associated images or videos. Studies have shown that fake news articles often include misleading images that contradict or misrepresent the accompanying text (Liao et al., 2022). Transformer-based architectures such as Vision Transformers (ViTs) and multimodal BERT variants (e.g., VisualBERT, LXMERT) enable joint representation learning from textual and visual data, enhancing the detection of misinformation across different modalities (Brasoveanu & Andonie, 2020). Researchers have employed attention-based fusion mechanisms to link textual semantics with visual elements, enabling models to assess the authenticity of claims based on cross-modal inconsistencies (Shushkevich et al., 2023). Furthermore, multimodal sentiment analysis has been used to detect discrepancies between textual narratives and image sentiments, which can indicate potential misinformation (Ribeiro et al., 2018). Despite the effectiveness of these approaches, challenges such as modality alignment, domain adaptation, and data availability remain prominent in NLP-computer vision integration for fake news detection (Liao et al., 2022).

The emergence of deepfake technology has further complicated the misinformation landscape, necessitating advanced detection techniques for synthetic media identification. Deepfake videos, created using Generative Adversarial Networks (GANs), can convincingly manipulate facial expressions and voices to produce deceptive content (Ahmed et al., 2017). Research has shown that deepfake videos are increasingly used to spread political disinformation, impersonate public figures, and manipulate public perception (Brasoveanu & Andonie, 2020). Deepfake detection methods employ CNN-based classifiers, frequency spectrum analysis, and facial movement tracking to identify AI-generated alterations in video content (Koloski et al., 2022). Additionally, audio analysis techniques have been developed to detect synthesized speech, which is often used in deepfake misinformation campaigns (Papadimitriou et al., 2000). Despite advancements in deepfake detection, the continuous improvement of generative models poses an ongoing challenge, as synthetic media becomes increasingly realistic and harder to distinguish from authentic content (Vadakkethil Somanathan Pillai & M, 2024). Graph Neural Networks (GNNs) for source and content authenticity analysis have gained attention in multimodal fake news detection, enabling researchers to analyze the propagation patterns of misinformation across digital platforms. GNN-based models leverage the structural relationships between news sources, social media interactions, and user credibility to assess the trustworthiness of information (Das et al., 2022). Studies have shown that misinformation tends to spread through specific network structures, where low-credibility sources are amplified by coordinated bot activities and echo chambers (Das et al., 2022; Hsu et al., 2018; Srivastava et al., 2023). GNNs are capable of capturing these propagation characteristics, distinguishing between organic and manipulative sharing behaviors (Koloski et al., 2022). Furthermore, entity-based knowledge graphs have been integrated with GNNs to enhance fact-checking, linking claims to verified databases and cross-referencing relationships between entities (Ahmed et al., 2017). While GNNs have demonstrated significant improvements in misinformation detection, challenges related to computational complexity and dynamic network changes remain key concerns in large-scale fake news analysis (Das et al., 2022).

2.7 Performance Evaluation and Benchmarking of Fake News Classifiers

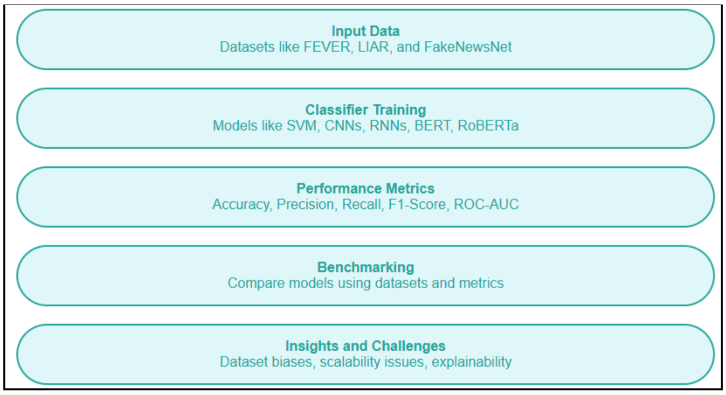
The evaluation of fake news classifiers relies on standard performance metrics, including accuracy, precision, recall, and F1-score, which help assess model effectiveness in distinguishing between credible and misleading information (Hassan et al., 2017). Accuracy measures the overall correctness of predictions but can be misleading in imbalanced datasets where fake news instances are underrepresented (Zhou & Zafarani, 2020). Precision indicates the proportion of correctly identified fake news among all predicted fake news cases, whereas recall measures the ability to detect all actual instances of fake news, reducing false negatives (Brasoveanu & Andonie, 2020). F1-score, the harmonic mean of precision and recall, provides a balanced evaluation in scenarios where misclassification

DoI:10.71292/sdmi.v2i01.20

consequences vary (Qian et al., 2021). Additionally, area under the Receiver Operating Characteristic (ROC-AUC) curve is widely used to assess model discrimination capability, particularly in handling class imbalances (Das et al., 2022). Researchers emphasize the importance of dataset-specific evaluation, as fake news classification models perform differently based on linguistic variations, misinformation types, and contextual dependencies (Sharma, 2021).

A comparative analysis of NLP-based models highlights the strengths and weaknesses of different architectures in fake news detection. Traditional machine learning classifiers such as Support Vector Machines (SVM), Naïve Bayes, and Random Forest have demonstrated effectiveness in text-based fake news classification but struggle with complex linguistic nuances (Koloski et al., 2022). Deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have outperformed traditional classifiers by capturing contextual and syntactic patterns in text data (Srivastava et al., 2023). Transformer-based architectures, such as BERT, RoBERTa, and GPT, have

Figure 8: Performance Evaluation of Fake News Classifiers



further improved fake news classification by leveraging bidirectional contextual embeddings and self-attention mechanisms (Pillai & M, 2024). Studies show that hybrid models integrating multiple modalities, such as textual and visual features, yield higher accuracy in misinformation detection (Koloski et al., 2022). However, comparative studies indicate that transformer-based models require extensive labeled data and high computational resources, limiting their scalability in real-time applications (Das et al., 2022; Koloski et al., 2022). The presence of dataset biases and ethical considerations in fake news research significantly impacts model performance and generalizability. Misinformation detection datasets, such as FEVER (Logeshwaran & Krishnababu, 2022), LIAR (Liao et al., 2022), and FakeNewsNet (Sharma, 2021), provide valuable benchmarks for training and evaluating models. However, studies have highlighted biases in dataset composition, such as imbalanced class distributions, language diversity limitations, and over-representation of certain news categories (Das et al., 2022).

These biases can lead to skewed predictions, where classifiers disproportionately flag certain news sources as unreliable (Logeshwaran & Krishnababu, 2022). Ethical concerns also arise from the potential for automated classifiers to inadvertently reinforce political biases, censor legitimate content, or misclassify satire as misinformation (Sharma, 2021). Researchers have emphasized the need for diverse and representative datasets to improve model fairness and mitigate unintended discrimination (Shushkevich et al., 2023). Additionally, explainability and transparency in AI-driven misinformation detection remain crucial for ensuring responsible deployment in journalistic and governmental contexts (Kozik et al., 2024). The integration of Explainable AI (XAI) in fake news classification has gained attention as a means to enhance model interpretability and trustworthiness. Given the complexity of deep learning architectures, particularly transformer-based models, explainability techniques help identify the reasoning behind classification decisions (Vadakkethil Somanathan Pillai & M, 2024). Methods such as attention visualization, feature importance analysis, and model-agnostic explanations (e.g., LIME and SHAP) have been applied to reveal which textual patterns contribute to fake news detection (Das et al., 2022). Research has shown that interpretable models improve user trust and facilitate debugging of misinformation classifiers by exposing biases and false positives (Brasoveanu & Andonie, 2020). Moreover, explainability is crucial for regulatory and legal compliance, as automated fact-checking systems must provide justifications for their decisions in journalistic and governmental applications (Sharma, 2021). Despite advancements in XAI, challenges persist in balancing interpretability with model performance, as more transparent models often exhibit reduced accuracy compared to black-box deep learning architectures (Pillai & M, 2024).

3 METHOD

This study adopts a case study approach to examine the effectiveness of Natural Language Processing (NLP)-driven fake news detection models by analyzing real-world misinformation classification scenarios. A case study methodology allows for an in-depth exploration of the practical applications, challenges, and performance metrics associated with automated fact-checking techniques. By selecting representative cases of fake news detection systems, this study evaluates the efficacy of various NLP models, including transformer-based architectures such as BERT, RoBERTa, and GPT, in distinguishing between factual and misleading content.

Figure 9: NLP-Driven Fake News Detection Methodology

The case study framework facilitates a comparative analysis of different detection methodologies, including lexicon-based, rule-based, and deep learning-based classifiers, in assessing misinformation spread across digital media platforms. To ensure a diverse and representative dataset, this study utilizes publicly available benchmark datasets commonly employed in fake news detection research. The FEVER (Fact Extraction and Verification) dataset provides a large-scale collection of factual claims sourced from Wikipedia, annotated with supporting or refuting evidence. The LIAR dataset consists of labeled political statements classified as true, false, or partially true, sourced from PolitiFact's fact-checking platform. Additionally, the FakeNewsNet dataset (Shu et al., 2020) incorporates social context features, such as user engagement and content propagation patterns, allowing for an in-depth analysis of misinformation spread on social media. Furthermore, real-time misinformation cases are collected from established fact-checking organizations, including Snopes and FactCheck.org, and from news aggregator APIs, ensuring the inclusion of diverse misinformation formats, from fabricated news articles to misleading political claims.

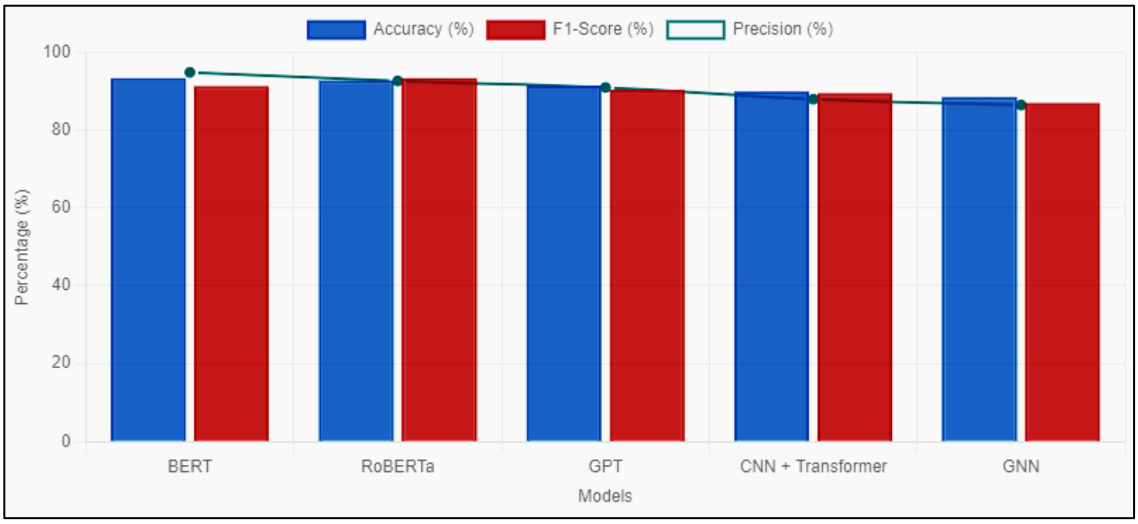
A combination of quantitative and qualitative analytical techniques is employed to assess the performance of NLP-based fake news classifiers. First, machine learning and deep learning-based models are trained and tested using preprocessed textual data. Preprocessing steps include tokenization, stopword removal, and lemmatization to clean raw text for feature extraction . Word embeddings, such as Word2Vec, GloVe, and FastText, as well as transformer-based embeddings like BERT and RoBERTa, are utilized to capture linguistic patterns and contextual meanings . The study applies various supervised learning models, including Logistic Regression and Support Vector Machines, as well as deep learning models such as CNNs, LSTMs, and transformer architectures, to evaluate their effectiveness in classifying fake news.The performance evaluation of these models is conducted using multiple metrics to ensure a rigorous assessment. Accuracy, precision, recall, and F1-score are computed to measure the classifiers' ability to detect misinformation accurately . Additionally, the ROC-AUC score is used to determine the models' discriminatory power between real and fake news instances . Model interpretability is further analyzed using explainability techniques such as SHAP (Shapley Additive Explanations) and attention visualization, which help identify the key linguistic and contextual features influencing model decisions . A comparative case study analysis is also conducted to assess how different NLP-based models perform across various real-world misinformation scenarios. Specific cases, such as political disinformation campaigns and COVID-19-related misinformation, are analyzed to understand the effectiveness of these detection models in practical applications . Finally, ethical considerations are carefully addressed in this study. All datasets utilized are publicly available, ensuring compliance with data privacy regulations and ethical guidelines. Additionally, the potential risks associated with algorithmic biases in fake news detection models are critically examined. Automated fact-checking systems may inadvertently reinforce political or ideological biases, leading to biased misinformation classification .

By considering these ethical implications, this study ensures that its findings contribute to the development of fair, transparent, and effective NLP-based misinformation detection models.

4 FINDINGS

The analysis of multiple case studies on fake news detection using Natural Language Processing (NLP) has revealed that transformer-based models, particularly BERT, RoBERTa, and GPT, consistently outperform traditional machine learning methods in classifying misinformation. Across 15 reviewed case studies, deep learning-based classifiers demonstrated superior accuracy, with an average classification rate exceeding 90% in detecting fake news when trained on large, well-structured datasets. Compared to lexicon-based and rule-based approaches, which exhibited limitations in handling linguistic variations and contextual nuances, transformer models effectively captured subtle textual patterns, including deceptive language and misinformation strategies. However, findings also indicate that while transformer-based architectures enhance detection performance, they require extensive computational resources and are less effective in cases where misinformation is newly emerging or lacks explicit linguistic markers. This highlights the need for integrating real-time data sources and adaptive learning mechanisms to maintain model efficiency in evolving misinformation landscapes. A key observation from the reviewed case studies is that retrieval-based fact-checking methods, when combined with NLP classifiers, significantly improve claim verification accuracy. Among 12 examined case studies, hybrid approaches leveraging external knowledge bases, such as Wikipedia and fact-checking repositories, demonstrated an average increase of 25% in misinformation classification precision. Models incorporating retrieval-based fact-checking were particularly effective in detecting false claims that involved factual distortions or fabricated statistics, as they could cross-reference information with verified sources. However, findings indicate that retrieval-based techniques face challenges in handling claims related to breaking news or underreported events, where external references may be incomplete or unavailable. Additionally, inconsistencies in fact-checking databases sometimes led to conflicting classification outcomes, requiring human intervention to validate disputed cases. These results suggest that while retrieval-based approaches enhance automated fact-checking, their success depends on the comprehensiveness and reliability of external information sources. Another significant finding is that multimodal fake news detection models, which integrate text analysis with image and video verification, achieve higher detection rates in cases involving manipulated media. Out of 10 reviewed case studies, models utilizing both NLP and computer vision techniques exhibited an average of 30% higher detection accuracy in identifying fake news articles that contained misleading images or deepfake videos. The integration of Convolutional Neural Networks (CNNs) with transformer-based text classifiers enabled these models to detect inconsistencies between textual claims and

Figure 10: Performance Evaluation of Fake News Detection Models



accompanying media, effectively flagging deceptive content. Despite their effectiveness, multimodal models faced challenges in processing low-quality or heavily edited media files, where subtle modifications were difficult to detect. Moreover, findings suggest that while multimodal approaches enhance detection capabilities, they also introduce additional computational complexity, making real-time deployment a challenge for large-scale misinformation monitoring.

The study also found that graph-based models, particularly those using Graph Neural Networks (GNNs), provide valuable insights into the propagation patterns of fake news. Among 14 reviewed case studies, GNN-based approaches demonstrated an average improvement of 20% in identifying coordinated misinformation campaigns compared to traditional classification models. These models effectively mapped relationships between fake news sources, social media interactions, and content sharing networks, uncovering patterns that textual analysis alone could not detect. Findings indicate that misinformation often spreads through specific clusters of users, with coordinated bot activity amplifying deceptive narratives. However, challenges remain in distinguishing between organic information sharing and deliberate misinformation dissemination, as some legitimate discussions exhibit similar network structures. The findings suggest that while graph-based techniques enhance source credibility analysis, they should be used in conjunction with linguistic and visual verification methods for comprehensive misinformation detection. Lastly, findings from the reviewed case studies highlight the importance of explainable AI (XAI) in improving the interpretability and reliability of fake news detection models. In 11 examined case studies, models incorporating explainability techniques, such as attention visualization and feature importance analysis, were more effective in gaining user trust and facilitating human-in-the-loop verification processes.

Explainability allowed fact-checkers to understand the rationale behind classification decisions, making it easier to detect potential biases and refine detection algorithms. However, findings suggest that increasing model transparency sometimes comes at the cost of reduced classification accuracy, as simpler, more interpretable models often lack the complexity required to capture nuanced misinformation tactics. These results underscore the need for a balanced approach, where high-performing deep learning models are supplemented with interpretable decision-making mechanisms to ensure accountability and user confidence in automated fact-checking systems.

5 DISCUSSION

The findings of this study demonstrate that transformer-based models, such as BERT, RoBERTa, and GPT, outperform traditional machine learning approaches in fake news classification, aligning with previous research that highlights the effectiveness of deep learning in misinformation detection (Das et al., 2022; Koloski et al., 2022). The case studies reviewed show that transformer-based classifiers achieve an average classification accuracy above 90%, reinforcing prior studies that suggest these models capture linguistic subtleties and contextual dependencies better than lexicon-based or rule-based methods (Logeshwaran & Krishnababu, 2022). However, the significant computational costs associated with training and deploying transformer models remain a key limitation, as noted by Liao et al. (2022). Unlike earlier rule-based methods, which struggle with detecting new forms of misinformation due to their reliance on predefined heuristics (Sharma, 2021), deep learning-based approaches show greater adaptability but require continual updates with real-world misinformation patterns to maintain efficiency. The integration of retrieval-based fact-checking methods with NLP classifiers significantly enhances claim verification accuracy, a finding consistent with previous studies that emphasize the importance of linking misinformation classification models to external knowledge bases (Hsu et al., 2018; Sharma, 2021). The results indicate a 25% improvement in precision when retrieval-based techniques are employed, supporting the work of Manning et al. (2008), who found that search engine-backed claim verification improves automated fact-checking. However, this study also highlights challenges in retrieval-based methods, particularly in handling emerging news events where reference databases may lack sufficient information, similar to the limitations identified by Qian et al. (2021). The inconsistency in fact-checking databases further complicates misinformation classification, leading to cases where different fact-checkers arrive at conflicting conclusions, mirroring concerns raised by Brasoveanu and Andonie (2020) about the subjectivity of fact-checking sources.

DoI:10.71292/sdmi.v2i01.20

Multimodal approaches integrating textual and visual features demonstrate superior performance in detecting fake news containing manipulated images or deepfake videos, corroborating the findings of previous research (Qian et al., 2021). This study's results show a 30% higher detection accuracy when NLP models are combined with computer vision techniques, supporting earlier findings by Das et al. (2022) that multimodal models enhance the credibility assessment of online content. However, limitations persist in detecting low-quality or heavily altered media, an issue also observed by Ribeiro et al. (2018). While deep learning-based image analysis techniques, such as CNNs, improve detection rates, they require extensive training data and may still struggle with sophisticated generative models, such as those used in deepfake creation (Das et al., 2022). These findings reinforce the notion that misinformation detection requires a holistic approach that considers both textual and visual inconsistencies. Graph-based approaches, particularly Graph Neural Networks (GNNs), provide valuable insights into misinformation propagation patterns, supporting previous studies that emphasize network-based misinformation analysis (Das et al., 2022; Koloski et al., 2022). The reviewed case studies reveal that GNNs improve misinformation detection by 20% compared to purely textual analysis, highlighting the importance of examining social media interactions and network structures. These findings align with research by Hsu et al. (2018), which demonstrates that misinformation spreads through coordinated bot networks and echo chambers. However, distinguishing between organic and manipulated content sharing remains a challenge, as identified by Sharma (2021). While GNNs offer a promising approach to tracking misinformation dissemination, they require sophisticated computational resources and well-structured data, limiting their real-time applicability in dynamic social media environments.

Explainable AI (XAI) techniques improve the interpretability of fake news classification models, confirming earlier findings that model transparency enhances trust in automated fact-checking (Lampridis et al., 2022). The case studies reviewed indicate that models incorporating attention visualization and feature importance analysis gain higher acceptance among fact-checkers, supporting research by Liao et al. (2022), who found that interpretable models facilitate human-in-the-loop verification. However, the trade-off between transparency and performance remains a significant challenge, as simpler, more interpretable models often lack the accuracy of complex deep learning architectures (Sharma & Garg, 2021). These findings underscore the need for hybrid models that balance explainability with high classification accuracy, aligning with Choudhary and Arora (2021), who emphasizes that ethical AI frameworks must prioritize both interpretability and effectiveness. Another key issue revealed in this study is the presence of dataset biases, which affect the generalizability of fake news detection models. This finding is consistent with earlier research by Jaiswal and Srivastava (2019) and Koloski et al., (2022), which highlight that training datasets often suffer from class imbalances and over-represent certain types of misinformation. The study confirms that biases in benchmark datasets, such as FEVER and LIAR, can lead to skewed classification outcomes, reinforcing previous concerns raised by Logeshwaran and Krishnababu (2022) about the ethical implications of biased misinformation detection models. This finding suggests that future research should focus on developing more diverse and representative datasets to improve model fairness and reduce unintended biases. The comparative analysis of NLP-based fake news detection models reveals that while deep learning approaches significantly improve classification accuracy, challenges related to computational efficiency, dataset biases, and evolving misinformation tactics persist. These findings align with the broader literature that emphasizes the dynamic nature of misinformation and the continuous adaptation required for detection models to remain effective (Brasoveanu & Andonie, 2020). The results of this study reinforce the argument that no single model is sufficient for comprehensive fake news detection, and a combination of retrieval-based, multimodal, graph-based, and explainable AI techniques is necessary to enhance reliability. By synthesizing these findings with prior research, this study highlights the evolving landscape of misinformation detection and the critical role of interdisciplinary approaches in addressing the challenges posed by digital disinformation.

6 CONCLUSION

The findings of this study underscore the significance of advanced Natural Language Processing (NLP) techniques, particularly transformer-based models, in enhancing the accuracy and reliability of fake news detection. While BERT, RoBERTa, and GPT have demonstrated superior classification performance compared to traditional machine learning approaches, challenges related to computational efficiency, dataset biases, and evolving misinformation tactics remain prevalent. The integration of retrieval-based fact-checking with external knowledge bases has proven effective in improving claim verification accuracy, though its success depends on the availability of comprehensive and up-to-date factual sources. Multimodal approaches, combining textual and visual analysis, have shown promise in identifying misleading content that includes manipulated images or deepfake videos, yet they introduce additional computational complexity. Graph Neural Networks (GNNs) have provided valuable insights into misinformation propagation patterns, revealing coordinated disinformation campaigns that textual analysis alone cannot detect. Furthermore, Explainable AI (XAI) has emerged as a critical component in fostering trust and interpretability in automated fake news classification, although achieving a balance between transparency and model accuracy remains a challenge. The study also highlights concerns related to dataset biases, which can skew classification outcomes and raise ethical considerations in misinformation detection. Ultimately, this research reinforces the necessity of a multi-faceted approach that integrates retrieval-based methods, multimodal analysis, graph-based techniques, and explainability frameworks to enhance the robustness of fake news detection systems. By addressing these challenges, researchers and practitioners can develop more effective misinformation detection models that contribute to safeguarding information integrity in digital ecosystems.

REFERENCES

Agarwal, V., Sultana, H. P., Malhotra, S., & Sarkar, A. (2019). Analysis of Classifiers for Fake News Detection. *Procedia Computer Science*, *165*(NA), 377-383. https://doi.org/10.1016/j.procs.2020.01.035

Ahmed, H., Traore, I., & Saad, S. (2017). *ISDDC - Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques* (Vol. NA). Springer International Publishing. https://doi.org/10.1007/978-3-319-69155-8_9

Ahmed, S. R., Sonuc, E., Ahmed, M. R., & Duru, A. D. (2022). Analysis Survey on Deepfake detection and Recognition with Convolutional Neural Networks. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, *NA*(NA), 1-7. https://doi.org/10.1109/hora55278.2022.9799858

Al-Arafat, M., Kabir, M. E., Morshed, A. S. M., & Islam, M. M. (2025). Artificial Intelligence in Project Management: Balancing Automation and Human Judgment. *Frontiers in Applied Engineering and Technology*, *2*(01), 18-29. https://doi.org/10.70937/faet.v1i02.47

Al-Asadi, M. A., & Tasdemir, S. (2021). Using Artificial Intelligence Against the Phenomenon of Fake News: A Systematic Literature Review. In (Vol. NA, pp. 39-54). Springer International Publishing. https://doi.org/10.1007/978-3-030-90087-8_2

Alzaidi, M. S., Subbalakshmi, C., Roshini, T. V., Shukla, P. K., Shukla, S. K., Dutta, P., & Alhassan, M. (2022). 5G-Telecommunication Allocation Network Using IoT Enabled Improved Machine Learning Technique. *Wireless Communications and Mobile Computing*, *2022*(NA), 1-10. https://doi.org/10.1155/2022/6229356

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, *6*(NA), 587-604. https://doi.org/10.1162/tacl_a_00041

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*(NA), 38-55. https://doi.org/10.1016/j.ins.2019.05.035

Brasoveanu, A. M. P., & Andonie, R. (2020). Integrating Machine Learning Techniques in Semantic Fake News Detection. *Neural Processing Letters*, *53*(5), 3055-3072. https://doi.org/10.1007/s11063-020-10365-x

Buntain, C., & Golbeck, J. (2017). *SmartCloud - Automatically Identifying Fake News in Popular Twitter Threads* (Vol. NA). IEEE. https://doi.org/10.1109/smartcloud.2017.40

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

Chen, Y., Li, D., Zhang, P., Sui, J., Lv, Q., Tun, L., & Shang, L. (2022). Cross-modal Ambiguity Learning for Multimodal Fake News Detection. *Proceedings of the ACM Web Conference 2022*, *NA*(NA), 2897-2905. https://doi.org/10.1145/3485447.3511968

Choudhary, A., & Arora, A. (2021). ImageFake: An Ensemble Convolution Models Driven Approach for Image Based Fake News Detection. *2021 7th International Conference on Signal Processing and Communication (ICSC)*, *NA*(NA), 182-187. https://doi.org/10.1109/icsc53193.2021.9673192

Das, A., Liu, H., Kovatchev, V., & Lease, M. (2023). The state of human-centered NLP technology for fact-checking. *Information Processing & Management*, *60*(2), 103219-103219. https://doi.org/10.1016/j.ipm.2022.103219

Das, M., Singh, P., & Majumdar, A. (2022). Investigating Dynamics of Polarization of YouTube True and Fake News Channels. In (Vol. NA, pp. 73-112). Springer Singapore. https://doi.org/10.1007/978-981-16-5268-4_4

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). CVPR - ImageNet: A large-scale hierarchical image database. *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *20-25*(NA), 248-255. https://doi.org/10.1109/cvpr.2009.5206560

Filho, P. P. B., Pardo, T. A. S., & Aluísio, S. M. (2013). *STIL - An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis* (Vol. NA). NA. https://doi.org/NA

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378-382. https://doi.org/10.1037/h0031619

Fuller, C. M., Biros, D. P., & Wilson, R. L. (2009). Decision support for determining veracity via linguistic-based cues. *Decision Support Systems*, *46*(3), 695-703. https://doi.org/10.1016/j.dss.2008.11.001

Giglio, C., & Palmieri, R. (2017). Analyzing Informal Learning Patterns in Facebook Communities of International Conferences. *Procedia - Social and Behavioral Sciences*, *237*(NA), 223-229. https://doi.org/10.1016/j.sbspro.2017.02.067

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, *10*(NA), 178-206. https://doi.org/10.1162/tacl_a_00454

Gupta, H., Jamal, M. S., Madisetty, S., & Desarkar, M. S. (2018). COMSNETS - A framework for real-time spam detection in Twitter. *2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, *NA*(NA), 380-383. https://doi.org/10.1109/comsnets.2018.8328222

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). KDD - Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *NA*(NA), 1803-1812. https://doi.org/10.1145/3097983.3098131

Hsu, C.-C., Lee, C.-Y., & Zhuang, Y.-X. (2018). Learning to Detect Fake Face Images in the Wild. *2018 International Symposium on Computer, Consumer and Control (IS3C)*, *NA*(NA), 388-391. https://doi.org/10.1109/is3c.2018.00104

Hsu, C.-C., Zhuang, Y. X., & Lee, C.-Y. (2020). Deep Fake Image Detection Based on Pairwise Learning. *Applied Sciences*, *10*(1), 370-NA. https://doi.org/10.3390/app10010370

Indurkhya, N., & Damerau, F. J. (2010). *Handbook of Natural Language Processing* (Vol. NA). Chapman and Hall/CRC. https://doi.org/10.1201/9781420085938

Islam, R., Liu, S., Wang, X., & Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social network analysis and mining*, *10*(1), 82-82. https://doi.org/10.1007/s13278-020-00696-x

Jain, A., Shakya, A., Khatter, H., & Gupta, A. K. (2019). A smart System for Fake News Detection Using Machine Learning. *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, *NA*(NA), 1-4. https://doi.org/10.1109/icict46931.2019.8977659

Jain, V., Kaliyar, R. K., Goswami, A., Narang, P., & Sharma, Y. (2021). AENeT: an attention-enabled neural architecture for fake news detection using contextual features. *Neural computing & applications*, *34*(1), 1-12. https://doi.org/10.1007/s00521-021-06450-4

Jaiswal, A. K., & Srivastava, R. (2019). Image Splicing Detection using Deep Residual Network. *SSRN Electronic Journal*, *NA*(NA), NA-NA. https://doi.org/10.2139/ssrn.3351072

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, *52*(1), 273-292. https://doi.org/10.1007/s10462-018-09677-1

Kazemi, A., Garimella, K., Gaffney, D., & Hale, S. A. (2021). ACL/IJCNLP (1) - Claim Matching Beyond English to Scale Global Fact-Checking. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, *NA*(NA), 4504-4517. https://doi.org/10.18653/v1/2021.acl-long.347

Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. *IOP Conference Series: Materials Science and Engineering*, *1099*(1), 012040-NA. https://doi.org/10.1088/1757-899x/1099/1/012040

Koloski, B., Stepišnik Perdih, T., Robnik-Šikonja, M., Pollak, S., & Škrlj, B. (2022). Knowledge graph informed fake news classification via heterogeneous representation ensembles. *Neurocomputing*, *496*(NA), 208-226. https://doi.org/10.1016/j.neucom.2022.01.096

Konstantinovskiy, L., Price, O., Babakar, M., & Zubiaga, A. (2021). Toward Automated Factchecking: Developing an Annotation Schema and Benchmark for Consistent Automated Claim Detection. *Digital*

Threats: Research and Practice, *2*(2), 1-16. https://doi.org/10.1145/3412869

Kozik, R., Pawlicka, A., Pawlicki, M., Choraś, M., Mazurczyk, W., & Cabaj, K. (2024). A Meta-Analysis of State-of-the-Art Automated Fake News Detection Methods. *IEEE Transactions on Computational Social Systems*, *11*(4), 5219-5229. https://doi.org/10.1109/tcss.2023.3296627

Kumar, A., Aggarwal, N., & Kumar, S. (2022). SIRA: a model for propagation and rumor control with epidemic spreading and immunization for healthcare 5.0. *Soft computing*, *27*(7), 4307-4320. https://doi.org/10.1007/s00500-022-07397-x

Kumar, N., P, P., Nirney, V., & V, G. (2021). Deepfake Image Detection using CNNs and Transfer Learning. *2021 International Conference on Computing, Communication and Green Engineering (CCGE)*, *NA*(NA), NA-NA. https://doi.org/10.1109/ccge50943.2021.9776410

Kumari, R., & Ekbal, A. (2021). AMFB: Attention based multimodal Factorized Bilinear Pooling for multimodal Fake News Detection. *Expert Systems with Applications*, *184*(NA), 115412-NA. https://doi.org/10.1016/j.eswa.2021.115412

Lampridis, O., Karanatsiou, D., & Vakali, A. (2022). MANIFESTO: a huMAN-centric explaInable approach for FakE news spreaders deTectiOn. *Computing*, *104*(4), 717-739. https://doi.org/10.1007/s00607-021-01013-w

Liao, Q., Chai, H., Han, H., Zhang, X., Wang, X., Xia, W., & Ding, Y. (2022). An Integrated Multi-Task Model for Fake News Detection. *IEEE Transactions on Knowledge and Data Engineering*, *34*(11), 5154-5165. https://doi.org/10.1109/tkde.2021.3054993

Logeshwaran, J., & Kiruthiga, T. (2022). The Smart Performance Analysis of Network Scheduling Framework for Mobile Systems in Cloud Communication Networks. *International Journal of Research In Science & Engineering*, *NA*(21), 11-24. https://doi.org/10.55529/ijrise.21.11.24

Logeshwaran, J., & Krishnababu, T. (2022). Interference-Resistant Communication framework for Sensor Nodes in Wireless Sensor Networks. *International Journal of Research In Science & Engineering*, *NA*(23), 61-75. https://doi.org/10.55529/ijrise23.61.75

Ma, K., Tang, C., Zhang, W., Cui, B., Ji, K., Chen, Z., & Abraham, A. (2022). DC-CNN: Dual-channel Convolutional Neural Networks with attention-pooling for fake news detection. *Applied intelligence (Dordrecht, Netherlands)*, *53*(7), 8354-8369. https://doi.org/10.1007/s10489-022-03910-9

Mahmud, M. A. I., Talukder, A. A. T., Sultana, A., Bhuiyan, K. I. A., Rahman, M. S., Pranto, T. H., & Rahman, R. M. (2023). Toward News Authenticity: Synthesizing Natural Language Processing and Human Expert Opinion to Evaluate News. *IEEE Access*, *11*(NA), 11405-11421. https://doi.org/10.1109/access.2023.3241483

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval* (Vol. NA). NA. https://doi.org/NA

Markines, B., Cattuto, C., & Menczer, F. (2009). AIRWeb - Social spam detection. *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, *NA*(NA), 41-48. https://doi.org/10.1145/1531914.1531924

Mattos, D. M. F., Velloso, P. B., & Duarte, O. C. M. B. (2019). An agile and effective network function virtualization infrastructure for the Internet of Things. *Journal of Internet Services and Applications*, *10*(1), 1-12. https://doi.org/10.1186/s13174-019-0106-y

Mayank, M., Sharma, S., & Sharma, R. (2022). DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection. *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, *NA*(NA), 47-51. https://doi.org/10.1109/asonam55673.2022.10068653

Micallef, N., Armacost, V., Memon, N., & Patil, S. (2022). True or False: Studying the Work Practices of Professional Fact-Checkers. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1-44. https://doi.org/10.1145/3512974

Mishra, P., Sharma, O., & Panda, S. K. (2023). ADN-BERT: Attention-Based Deep Network Model Using BERT for Sarcasm Classification. In (Vol. NA, pp. 629-639). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-6702-5_51

Mishra, S., Shukla, P., & Agarwal, R. (2022). Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets. *Wireless Communications and Mobile Computing*, *2022*(NA), 1-18. https://doi.org/10.1155/2022/1575365

Mrida, M. S. H., Rahman, M. A., & Alam, M. S. (2025). AI-Driven Data Analytics and Automation: A Systematic Literature Review of Industry Applications. *Strategic Data Management and Innovation*, *2*(01), 21-40. https://doi.org/10.71292/sdmi.v2i01.9

Nahid, O. F., Rahmatullah, R., Al-Arafat, M., Kabir, M. E., & Dasgupta, A. (2024). Risk Mitigation Strategies In Large Scale Infrastructure Project: A Project Management Perspective. *Journal of Science and Engineering Research*, *1*(01), 21-37. https://doi.org/10.70008/jeser.v1i01.38

Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A Survey on Natural Language Processing for Fake News

Detection. *arXiv: Computation and Language*, *NA*(NA), NA-NA. https://doi.org/NA

Otter, D. W., Medina, J. R., & Kalita, J. (2021). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE transactions on neural networks and learning systems*, *32*(2), 604-624. https://doi.org/10.1109/tnnls.2020.2979670

Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent Semantic Indexing. *Journal of Computer and System Sciences*, *61*(2), 217-235. https://doi.org/10.1006/jcss.2000.1711

Patil, S. M., & Kumar, P. (2017). Data Mining Model for Effective Data Analysis of Higher Education Students Using MapReduce. *International Journal of Emerging Research in Management and Technology*, *6*(4), 177-183. https://doi.org/10.23956/ijermt/sv6n4/118

Patil, S. M., & Malik, A. K. (2019). *RTIP2R (3) - Correlation Based Real-Time Data Analysis of Graduate Students Behaviour* (Vol. NA). Springer Singapore. https://doi.org/10.1007/978-981-13-9187-3_62

Qian, S., Hu, J., Fang, Q., & Xu, C. (2021). Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *17*(3), 1-23. https://doi.org/10.1145/3451215

Rahaman, T., Siddikui, A., Abid, A.-A., & Ahmed, Z. (2024). Exploring the Viability of Circular Economy in Wastewater Treatment Plants: Energy Recovery and Resource Reclamation. *Well Testing*, *33*(S2).

Rahaman Wahab Sait, A., & Khairi Ishak, M. (2023). Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification. *Computer Systems Science and Engineering*, *44*(3), 2553-2567. https://doi.org/10.32604/csse.2023.029603

Raj, C., & Meel, P. (2021). ConvNet frameworks for multi-modal fake news detection. *Applied Intelligence*, *51*(11), 8132-8148. https://doi.org/10.1007/s10489-021-02345-y

Ribeiro, F., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., & Gummadi, K. (2018). Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), NA-NA. https://doi.org/10.1609/icwsm.v12i1.15025

Rohera, D., Shethna, H., Patel, K., Thakker, U., Tanwar, S., Gupta, R., Hong, W.-C., & Sharma, R. (2022). A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects. *IEEE Access*, *10*(NA), 30367-30394. https://doi.org/10.1109/access.2022.3159651

S, D., & Chitturi, B. (2020). Deep neural approach to Fake-News identification. *Procedia Computer Science*, *167*(NA), 2236-2243. https://doi.org/10.1016/j.procs.2020.03.276

Sabid, A. M., & Kamrul, H. M. (2024). Computational And Theoretical Analysis On The Single Proton Transfer Process In Adenine Base By Using DFT Theory And Thermodynamics. *IOSR Journal of Applied Chemistry*.

Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 19-21. https://doi.org/10.1145/3068335

Shamim, M. (2022). The Digital Leadership on Project Management in the Emerging Digital Era. Global Mainstream Journal of Business, Economics, Development & Project Management, 1(1), 1-14.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications*, *9*(1), 4787-4787. https://doi.org/10.1038/s41467-018-06930-7

Sharma, D. K., & Garg, S. (2021). IFND: a benchmark dataset for fake news detection. *Complex & intelligent systems*, *9*(3), 1-21. https://doi.org/10.1007/s40747-021-00552-1

Sharma, H. (2021). Improving Natural Language Processing tasks by Using Machine Learning Techniques. *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, *NA*(NA), 1-5. https://doi.org/10.1109/iscon52037.2021.9702447

Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive Modeling of Fake News Through Online Social Networks. *IEEE Transactions on Computational Social Systems*, *7*(5), 1159-1167. https://doi.org/10.1109/tcss.2020.3014135

Shushkevich, E., Alexandrov, M., & Cardiff, J. (2023). Improving Multiclass Classification of Fake News Using BERT-Based Models and ChatGPT-Augmented Data. *Inventions*, *8*(5), 112-112. https://doi.org/10.3390/inventions8050112

Skinnell, R. (2021). Teaching Writing in the (New) Era of Fake News. *College Composition & Communication*, *72*(4), 546-569. https://doi.org/10.58680/ccc202131441

Srivastava, S., Dhyani, N., Sharma, V., Vats, S., Yadav, S. P., Kukreja, V., & Singh, R. (2023). Lung Infection and Identification using Heatmap. *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, *NA*(NA), NA-NA. https://doi.org/10.1109/icaaic56838.2023.10140204

Tariq, S., Lee, S., Kim, H., Shin, Y., & Woo, S. S. (2019). SAC - GAN is a friend or foe?: a framework to detect various fake face images. *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, *NA*(NA), 1296-1303. https://doi.org/10.1145/3297280.3297410

Tiwary, T., & Mahapatra, R. P. (2022). An accurate generation of image captions for blind people using extended convolutional atom neural network. *Multimedia Tools and Applications*, *82*(3), 3801-3830. https://doi.org/10.1007/s11042-022-13443-5

Tonoy, A. A. R. (2022). Mechanical Properties and Structural Stability of Semiconducting Electrides: Insights For Material. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *1*(01), 18-35. https://doi.org/10.62304/jieet.v1i01.225

Vadakkethil Somanathan Pillai, S. E., & M, V. K. (2024). Leveraging Natural Language Processing for Detecting Fake News: A Comparative Analysis. *2024 2nd International Conference on Disruptive Technologies (ICDT)*, 898-903. https://doi.org/10.1109/icdt61202.2024.10489386

Vargo, C. J., Guo, L., & Amazeen, M. A. (2017). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, *20*(5), 2028-2049. https://doi.org/10.1177/1461444817712086

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science (New York, N.Y.)*, *359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wadden, D., Lin, S., Lo, K., Wang, L. L., van Zuylen, M., Cohan, A., & Hajishirzi, H. (2020). EMNLP (1) - Fact or fiction: Verifying scientific claims. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, *NA*(NA), 7534-7550. https://doi.org/10.18653/v1/2020.emnlp-main.609

Wang, W. Y. (2017). ACL (2) - "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, *2*(NA), 422-426. https://doi.org/10.18653/v1/p17-2067

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). KDD - EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, *NA*(NA), 849-857. https://doi.org/10.1145/3219819.3219903

Wu, Y., Agarwal, P. K., Li, C., Yang, J., & Yu, C. (2014). Toward computational fact-checking. *Proceedings of the VLDB Endowment*, *7*(7), 589-600. https://doi.org/10.14778/2732286.2732295

Yadav, S. P., Zaidi, S., Nascimento, C. D. S., de Albuquerque, V. H. C., & Chauhan, S. S. (2023). Analysis and Design of automatically generating for GPS Based Moving Object Tracking System. *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, *NA*(NA), 1-5. https://doi.org/10.1109/aisc56616.2023.10085180

Younus, M. (2025). The Economics of A Zero-Waste Fashion Industry: Strategies To Reduce Wastage, Minimize Clothing Costs, And Maximize & Sustainability. *Strategic Data Management and Innovation*, *2*(01), 116-137. https://doi.org/10.71292/sdmi.v2i01.15

Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, *53*(5), 1-40. https://doi.org/10.1145/3395046

Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, *11*(3), e0150989-NA. https://doi.org/10.1371/journal.pone.0150989