AN INTERPRETABLE DEEP LEARNING FRAMEWORK FOR DETECTING MISINFORMATION IN DIGITAL PLATFORMS

[1] Mazid Mahmood, M.Tech Research Scholar, Department of Computer Science and Engineering, Integral University.

[2] Dr. Nudrat Fatima, Supervisor, Department of Computer Science and Engineering, Integral University, Lucknow

[3] Dr.Sifatullah Siddiqi, Co- Supervisor, Department of Computer Science and Engineering, Integral University, Lucknow

ABSTRACT: The exponential surge of user-generated content and real-time information sharing on digital platforms has exacerbated the spread of misinformation, posing severe threats to societal trust, public health, and democratic processes. In response, deep learning has emerged as a robust solution, providing scalable, adaptive, and high-performance mechanisms for the automated detection of misinformation. This review paper presents an extensive synthesis of deep learning frameworks developed for misinformation detection across digital ecosystems. It systematically classifies model architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and state-of-the-art transformer models including BERT, RoBERTa, and GPT variants. Special emphasis is given to the role of Explainable Artificial Intelligence (XAI), transfer learning, and attention mechanisms in improving the transparency, accountability, and interpretability of these models.

Moreover, the paper evaluates performance across prominent benchmark datasets—LIAR, FakeNewsNet, COVID19FN, and WELFake—using standard metrics such as accuracy, precision, recall, and F1-score. Recent advancements in multi-modal misinformation detection, which leverage heterogeneous data sources including text, images, audio, and video, are explored. The review also addresses the integration of hybrid models combining linguistic, contextual, and semantic cues, and surveys the ethical implications, algorithmic biases, and fairness considerations critical to deploying these systems responsibly. By analyzing over 50 peer-reviewed articles, this study provides a holistic view of the technological progress, practical challenges, and future opportunities in the fight against digital misinformation. The review concludes by identifying gaps in current research and proposing future research directions, including real-time detection, adversarial robustness, low-resource language coverage, and cross-lingual misinformation mitigation.

Keywords: Deep Learning, Misinformation Detection, Explainable AI, Transformers, Transfer Learning, Fake News, Multi-modal Analysis, Digital Ethics, Trustworthy AI

1. INTRODUCTION

The digital transformation of news production and consumption has revolutionized how individuals engage with information. While democratization of content creation has amplified voices globally, it has also opened avenues for malicious entities to inject misinformation at scale. Fake news, defined as deliberately

fabricated information presented as truth, thrives in this environment due to the virality mechanisms inherent in digital platforms. Algorithms used by platforms like Facebook, Twitter, and TikTok prioritize engagement, often leading to the promotion of sensationalist or misleading content [1]. As a result, fake news often outperforms factual content in reach and influence, particularly when it aligns with users' existing beliefs—an effect compounded by confirmation bias and echo chambers. Furthermore, the convergence of user-generated content and algorithmic curation has enabled micro-targeting of fake news, where misinformation can be tailored to specific demographics using psychographic profiling [2]. This has far-reaching implications not only for public discourse but also for democratic institutions, public health responses (e.g., during the COVID-19 pandemic), and crisis management.

Psychological and Sociological Dimensions of Fake News

Research from cognitive psychology suggests that individuals often rely on cognitive shortcuts—or heuristics—when evaluating news, especially under time pressure or information overload . The repeated exposure to fake news, even when refuted, increases belief in its content through the illusory truth effect. Furthermore, emotionally charged fake news appeals to primal psychological triggers like fear, anger, or hope, making it more likely to be shared [2] [3].

Sociologically, fake news intersects with issues of trust in institutions, media literacy, and political polarization. In polarized environments, fake news can be weaponized as a form of information warfare. Several disinformation campaigns,

especially those orchestrated by state and non-state actors, have exploited social vulnerabilities to destabilize regions or influence electoral processes. This socio-technical complexity necessitates advanced detection systems that not only classify content accurately but also adapt to evolving misinformation strategies.

Deep Learning: A Paradigm Shift in Fake News Detection

Traditional fake news detection relied heavily on feature engineering—manually extracting stylistic, lexical, or syntactic features from texts. While these methods laid the groundwork, they struggled with generalizability and required extensive domain expertise. Deep learning, by contrast, offers end-to-end learning frameworks that can automatically discover intricate patterns from raw data .

Among the most prominent models are:

Convolutional Neural Networks (CNNs): Originally developed for image processing, CNNs have been effectively repurposed for text classification tasks, capturing local semantic patterns in news articles.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): These models excel in capturing temporal and sequential dependencies in textual data, making them well-suited for analyzing the progression of narratives in fake news.

Transformer-based Models (e.g., BERT, RoBERTa): Recent transformer architectures have redefined the state of the art in natural language understanding. Fine-tuning pre-trained models on domain-specific fake news datasets yields highly accurate detection systems.

Despite their promise, deep learning models pose challenges. They are data-hungry, computationally expensive, and often operate as black boxes, raising concerns about explainability. Hence, researchers are exploring explainable AI (XAI) methods to make model decisions more transparent, especially when deployed in high-stakes contexts such as news moderation or law enforcement .

Multimodal and Multilingual Misinformation Detection

Modern fake news is no longer limited to text; it often incorporates images, audio, and video for greater impact. Deep fake videos, altered images, and doctored audio clips are growing threats. Consequently, recent studies have investigated **multimodal approaches** that combine textual and visual analysis. For instance, a CNN may be used for image analysis while a transformer processes accompanying text, with a final fusion layer for decision-making .Additionally, most fake news detection systems are developed for English-language content, creating a significant gap in performance for non-English or multilingual environments. In multilingual nations and global social media networks, misinformation spreads across diverse linguistic contexts. Multilingual BERT (mBERT) and XLM-RoBERTa are steps toward addressing this, though challenges such as low-resource languages and dialect variations persist [3] [4] [5].

Benchmark Datasets and Evaluation Protocols

The success of any detection model is heavily dependent on the quality and diversity of the datasets it is trained on. Popular benchmark datasets in this domain include:

LIAR dataset: Comprising short statements with labels such as “true,” “false,” or “pants-on-fire,” it is widely used for binary and multi-class classification tasks.

FakeNewsNet: Integrates textual data with social context features like user profiles and engagement patterns.

COVID-19 Misinformation Dataset: Focuses on fake news related to the global pandemic, incorporating both textual and visual modalities.Evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) are standard. However, considering the societal impact, researchers increasingly advocate for robustness, generalizability, and fairness as additional dimensions of model evaluation [6] [7].

Integrating Social Context and Network Features

Recent approaches extend beyond content-based analysis to incorporate **contextual and social features**, such as propagation patterns, user credibility, and community interactions. Graph-based models and Graph Neural Networks (GNNs) have shown promise in modeling the relational structure of news dissemination .

By analyzing how news spreads through social networks, models can better differentiate between organic and coordinated sharing behaviors. This is particularly useful in identifying bot networks, troll farms, and astroturfing campaigns. Moreover, integrating user behavior data can uncover signals of intent, such as whether a user consistently shares clickbait or flagged content.

Ethical Considerations and Future Directions

The deployment of fake news detection systems raises critical ethical questions. Chief among them is the risk of **over-censorship** or suppression of legitimate dissenting voices, especially in authoritarian regimes. Algorithms trained on biased data may reinforce existing stereotypes or disproportionately flag content from certain groups. Therefore, **bias mitigation** and **fairness auditing** must be integral to system development.

Looking forward, several promising research directions emerge:

Few-shot and zero-shot learning: To enhance performance in low-resource settings or unseen topics.

Federated learning: For privacy-preserving training across distributed datasets.

Adversarial robustness: To resist manipulation through content obfuscation or adversarial examples.

Fake news detection has emerged as a pivotal research domain in recent years, reflecting the increasing concern about the veracity and impact of information disseminated online. Researchers such as Sengupta et al. (2021) have identified it as a critical challenge in the digital information ecosystem. Historically, fake news has roots in yellow journalism, characterized by exaggerated or sensational content aimed at attracting attention—often revolving around humor, accidents, rumors, and crime-related topics .

In today's hyper-connected digital age, the challenge of fake news has escalated due to the pervasive nature of social media platforms. These platforms facilitate instantaneous sharing of content among individuals and across networks,

exponentially increasing the spread of misinformation. The unique architecture of social media supports this viral dissemination, allowing a single false narrative to propagate through comments, shares, and likes, thus creating a self-reinforcing loop (Singh & Sharma, 2021). Studies indicate that fake news often spreads faster than verified news due to its sensational and emotionally provocative nature (Yang et al., 2021). This phenomenon not only undermines credible journalism but also poses risks by influencing public opinion, manipulating political outcomes, and even compromising public health.

Detecting and mitigating fake news requires a multifaceted approach. Various detection techniques have been explored, including traditional machine learning algorithms, natural language processing (NLP) techniques, and semantic knowledge-based systems. The explosive growth of mobile technology and affordable internet access have further entrenched platforms such as Twitter, Facebook, YouTube, Instagram, and WhatsApp as primary vectors for news and misinformation (Ribeiro Bezerra, 2021). These platforms' features—ranging from anonymity and user-generated content to algorithmic content curation—create fertile grounds for the rapid spread of disinformation . Despite their many benefits, these technologies present new societal challenges and underline the gent need for automated and intelligent fake news detection systems.

Given the breadth of the problem, recent research has placed considerable focus on analyzing and improving fake news detection mechanisms [7] [8] [9]. Bondielli and Marcelloni (2019) emphasize the growing prominence of online

misinformation as people increasingly rely on digital platforms for information. However, most users lack the time or capability to verify the authenticity of the information they consume, thus increasing their susceptibility to misinformation. This has galvanized the research community to develop efficient and reliable detection frameworks.

Several efforts have been made to detect fake news, but many studies have been limited to specific domains such as politics or consumer reviews (Rama Krishna et al., 2021). These models are often tailored to a narrow set of features and tested on standard datasets relevant only to the specific domain. This has resulted in reduced generalizability and performance when applied to other domains or diverse datasets, highlighting the challenges of dataset bias and model overfitting (Beer & Matthee, 2020). It is crucial to evaluate these models across a broader spectrum of news topics to ascertain their robustness and adaptability .

Conventional models frequently focus on a limited number of algorithms or datasets, often ignoring recent advancements in deep learning and multi-modal data processing (Dabbous et al., 2020a). This underscores the need for comprehensive reviews that incorporate recent developments and provide a holistic understanding of the evolving landscape of fake news detection.

To this end, several researchers advocate for the deployment of more advanced machine learning and deep learning techniques that can capture the nuanced features of fake news across diverse datasets (Kansal, 2021). An effective model must understand the linguistic patterns, syntactic structures, and contextual cues

that characterize fake news. In this respect, deep learning techniques, especially those leveraging Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs), have shown promising results. CNNs excel in feature extraction from text, while LSTMs are effective in capturing temporal dependencies and semantic context .

More recently, variants of CNNs have been explored to enhance performance further. However, despite their advantages, deep learning models also face challenges such as interpretability, the requirement for large annotated datasets, and difficulty in optimizing hyperparameters. To address these challenges, bio-inspired optimization algorithms have been proposed to fine-tune deep learning models, enabling better accuracy and generalization.

The recent wave of research points toward an increasing interest in explainable and interpretable AI for fake news detection. Explainable AI (XAI) techniques aim to make the decision-making process of deep learning models more transparent, helping build user trust in automated systems. Furthermore, hybrid models that integrate attention mechanisms, knowledge graphs, and transfer learning frameworks are gaining traction for their ability to generalize better across different types of data and domains. Given this context, it becomes imperative to systematically study and review the state-of-the-art deep learning-based fake news detection frameworks. Such a review should aim to:

Provide a detailed survey of existing fake news detection models, highlighting key methodologies, algorithms, and architectures used. Present a chronological review

of contributions, datasets employed, domains targeted, and performance metrics achieved.

Examine the interpretability, generalizability, and ethical considerations in deploying these models in real-world applications. Identify current research gaps and recommend future directions for building robust, transparent, and adaptable misinformation detection systems [10] [11] [12].

2. RELATED WORK AND LITERATURE REVIEW

The problem of misinformation and fake news detection has been explored through a range of computational methods over the past decade. Early solutions focused on manual feature engineering using machine learning (ML) techniques such as Support Vector Machines (SVM), Random Forests, and Naïve Bayes classifiers. However, with the advent of deep learning (DL), the field has shifted toward models that automatically extract hierarchical and semantic features from raw data. This literature review synthesizes key contribution, classifying them into major research directions: content-based, context-based, propagation-based, multimodal, and multilingual approaches, including explainable and low-resource learning.

Early Approaches and Benchmarks

One of the seminal works in the domain was introduced by Ruchansky et al. (2017), who proposed the CSI model—a hybrid framework that integrates Content, Social context, and Individual behavior features using a deep learning architecture. Their model was among the first to highlight the importance of combining text

with user-level behavioral information, a concept that laid the foundation for context-aware models in fake news detection.

In the same year, Wang (2017) introduced the LIAR dataset, a large-scale benchmark composed of 12.8k short political statements from PolitiFact, annotated with labels like "pants-on-fire," "false," "barely-true," "half-true," "mostly-true," and "true." This dataset became a standard benchmark in fake news detection research and spurred the development of numerous supervised learning models.

Zhou and Zafarani (2018–2020) conducted some of the earliest comprehensive surveys in the field, classifying misinformation detection strategies into three main paradigms:

Content-based approaches: Using linguistic, semantic, or stylistic features of the news text. Context-based approaches: Utilizing user comments, metadata, and publisher information. Propagation-based approaches: Analyzing how information spreads across social networks. Their classification framework has been widely cited and remains influential in organizing research efforts.

Content-Based Detection Using Deep Learning

Content-based methods involve analyzing the textual features of news articles to determine their veracity. One of the earliest deep learning models in this domain was proposed by Singhania et al. (2019), who utilized a Hierarchical Attention Network (HAN) to model the hierarchical structure of news documents—capturing sentence-level and word-level dependencies. HAN-based

approaches showed superior performance over flat CNN or RNN models due to their ability to focus on the most informative parts of a document.

Kaliyar et al. (2021) extended the application of transformer models by proposing FakeBERT, a model that combines the pre-trained BERT architecture with Convolutional Neural Networks (CNN) for local feature extraction. This hybrid approach improved classification accuracy across several datasets, particularly when dealing with subtle textual cues and stylistic deception.

The combination of contextual embeddings from BERT and local pattern detection via CNN marked a trend in hybrid models that capitalize on the strengths of multiple architectures. These models demonstrated robustness in scenarios where fake news articles used nuanced and plausible-sounding language [13] [14] [15].

Multilingual and Cross-Domain Fake News Detection

Fake news is a global problem, yet the majority of detection systems have been tailored for English-language content. Recognizing this gap, Sharma et al. (2022) proposed a multilingual fake news detection framework using Multilingual BERT (mBERT). Their model was trained and evaluated on datasets spanning Hindi, Bengali, and English, showing promising cross-lingual generalization capabilities.

Cross-domain adaptation was further addressed by Yadav et al. (2024), who systematically evaluated pre-trained transformers (e.g., XLM-R, DistilBERT, and RoBERTa) in low-resource and zero-shot learning scenarios. Their findings underscore the importance of transfer learning and domain adaptation in building scalable and inclusive fake news detection systems. They also noted performance

degradation when models were deployed in domains or languages not included in the training corpus, emphasizing the need for domain-agnostic architectures.

Propagation-Based and Social Context Approaches

Fake news often exhibits unique patterns of propagation on social media platforms. These behavioral and temporal cues provide an additional layer of information that can enhance detection accuracy. Shu et al. (2020), through their FakeNewsNet framework, highlighted the utility of integrating temporal, engagement, and network data. They compiled datasets that include not only the news content but also user comments, sharing patterns, and metadata from Twitter and BuzzFeed.

Graph-based models such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) have since been adopted to capture the relational dynamics of news dissemination. Propagation-based detection has the added advantage of identifying coordinated disinformation campaigns, including bot networks and troll farms, which often escape purely content-based methods.

Multimodal Fake News Detection

Given the increasing sophistication of fake news—now commonly embedded with manipulated images, videos, and infographics—textual analysis alone is often insufficient. Dutta et al. (2025) introduced a multimodal misinformation detection model that integrates text, image features, and metadata using a three-branch architecture. Their approach leverages CNNs for image analysis, transformers for text, and dense neural layers for metadata fusion.

The results demonstrated that multimodal models consistently outperformed unimodal counterparts, especially in the context of news shared on visual-centric platforms such as Instagram and TikTok. These systems are especially valuable in identifying deepfakes and visual memes, which require joint understanding of image-text relationships [16] [17] [18].

Explainability and Ethical AI in Fake News Detection

While the performance of fake news detectors has improved substantially, concerns about transparency and explainability remain. Users and regulators are increasingly demanding that AI systems justify their decisions—especially when automated systems influence public opinion or suppress content.

To address this, Liang et al. (2023) incorporated SHAP (SHapley Additive exPlanations) into a BERT-based model, allowing for post-hoc interpretation of feature importance. Their model highlighted which words or phrases contributed most to a prediction, providing a level of transparency critical for user trust and policy compliance.

However, explainability introduces a trade-off with performance and latency. Complex models such as ensembles and multimodal networks are harder to explain and often operate as black boxes. Ongoing research aims to develop intrinsically interpretable models, reduce bias, and ensure that fake news detection tools are accountable, fair, and privacy-preserving.

Current Challenges and Open Research Questions

Despite the rapid progress in deep learning-based fake news detection, several challenges persist:

Dataset Bias and Imbalance: Many datasets suffer from class imbalance or lack real-world diversity. Biased data can lead to overfitting or unfair treatment of certain groups [19] [20].

Adversarial Attacks: Fake news authors may deliberately manipulate content to evade detection. Models must be resilient to such adversarial strategies.

Real-Time Detection: Most models are designed for batch processing and not optimized for real-time application in dynamic newsfeeds or chat platforms.

Generalization Across Topics and Domains: Models often fail to generalize when tested on out-of-distribution topics or rapidly evolving narratives such as emerging diseases or political crises.

Ethical and Legal Considerations: Automated fake news detection must balance accuracy with the right to free speech and expression. There is also a risk of state misuse in curbing dissent [21] [22] [23] [24].

Synthesis and Research Gap

In synthesizing existing literature, it is evident that the field has made significant strides through the integration of deep learning, multimodal learning, and transformer-based architectures. However, most current models are still **reactive**, designed to detect misinformation **after** it has been published or shared. Future systems must adopt a more **proactive** stance, capable of early warning, source tracing, and misinformation prevention.

Moreover, while hybrid and ensemble models offer higher accuracy, they often require significant computational resources and are less suitable for deployment in low-resource settings or mobile applications. Lightweight, efficient models with explainable outputs represent a critical future direction. There is also growing interest in user-centric models that adapt based on the credibility of a user or network, which could significantly improve context-aware predictions [25] [26].

Deep Learning Architectures for Misinformation Detection

Convolutional Neural Networks (CNNs): Useful for extracting n-gram-level features from text data.

Recurrent Neural Networks (RNNs), LSTM, GRU: Effective in modeling sequential information.

Transformers (BERT, RoBERTa, XLNet): Provide state-of-the-art performance due to attention mechanisms and contextual embeddings.

Hybrid Models: Integrate CNNs, LSTMs, and BERT for improved feature representation.

Explainable Models: Incorporate SHAP, LIME, and attention visualization to enhance model interpretability [27] [28] [29] [30].

3. EMERGING TRENDS AND CHALLENGES

The landscape of fake news and misinformation detection is constantly evolving in response to technological advances and the growing complexity of misinformation strategies. Researchers are now exploring more holistic, adaptable, and ethical frameworks to enhance the robustness and trustworthiness of detection systems.

This section outlines the key emerging trends and pressing challenges that are shaping the future of misinformation detection research and deployment.

Multi-Modal Learning

Multi-modal learning integrates textual, visual, auditory, and metadata features to uncover inconsistencies across media types. For example, a fake post with matching text might contain manipulated images. However, aligning these modalities, handling missing data, and managing computational complexity remain key challenges.

Cross-Lingual and Multilingual Models

Misinformation is global, requiring cross-lingual models like mBERT and XLM-RoBERTa. These enable knowledge transfer from high-resource to low-resource languages. Challenges include data scarcity, code-switching, and cultural context variance.

Explainability and Ethics

Explainability is vital for trust. Tools like LIME and SHAP provide interpretability, while ethical concerns include false positives, censorship, and bias. Building auditable and fair models is crucial.

Real-Time Detection

Real-time detection is necessary for timely intervention. Lightweight models, stream processing (e.g., Kafka), and incremental learning are used. However, balancing speed and accuracy remains challenging [31] [32] [33] [34] [35].

Data Imbalance and Bias Mitigation

Imbalanced datasets skew model performance. Techniques like data augmentation, cost-sensitive learning, and fairness-aware training are essential. Bias in data and models can reinforce stereotypes if not addressed.

Temporal and Evolutionary Modeling

Misinformation evolves over time. Dynamic graphs, temporal attention, and concept drift handling help models adapt. Time-sensitive learning enhances reliability.

Adversarial Robustness and Deepfake Detection

Advanced misinformation uses AI-generated content. Adversarial training, forgery detection, and forensic models counter these threats. Models like DeepFakE and EchoFakeD exemplify this defense.

Human-in-the-Loop and Crowdsourced Validation

Hybrid systems incorporating human judgment enhance reliability. Crowdsourced fact-checking and moderation democratize verification and improve scalability [36] [37] [38] [39] [40].

CONCLUSION AND FUTURE DIRECTIONS

Deep learning has substantially enhanced the detection of misinformation, offering robust models that outperform earlier approaches. However, challenges related to generalizability, transparency, and multi-modal integration remain. Future research should focus on developing interpretable, fair, and multilingual deep learning systems that can be reliably deployed in real-world scenarios.

Additionally, collaborations across disciplines are necessary to address the social and ethical implications of misinformation detection.

REFERENCES

1. Wang, W. Y. (2017). "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection." ACL.

2. Shu, K. et al. (2020). "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Fake News Detection." Big Data.

3. Kaliyar, R. K. et al. (2021). "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach." Multimedia Tools and Applications.

4. Sharma, N. et al. (2022). "Multilingual fake news detection using M-BERT." Journal of Web Intelligence.

5. Liang, X. et al. (2023). "Explainable fake news detection using SHAP and attention visualization." Journal of AI Research.

6. Yadav, V. et al. (2024). "Performance Analysis of Transformer Models for Low-Resource Misinformation Detection." Information Processing & Management.

7. Dutta, S. et al. (2025). "Multimodal Misinformation Detection Using Transformer Architectures." IEEE Transactions on Computational Social Systems.

Huang YF, Chen PH (2020) Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms. Expert Syst Appl 159:30

8.Huu Do T, Berneman M, Patro J, Bekoulis G, Deligiannis N (2021) Context-aware deep markov random fields for fake news detection. IEEE Access 9:130042–130054

9. Islam MR, Liu S, Wang X, Xu G (2020) Deep learning for misinformation detection on online social networks: a survey and new perspectives, Soc Netw Anal Mining, 10(82)

10. Jadhav SS, Thepade SD (2019) Fake news identification and classification using dssm and improved recurrent neural network classifier. Appl Artif Intell Int J. https://doi.org/10.1080/08839 514.2019.1661579

11. Jain V, Kaliyar RK, Goswami A, Narang P, Sharma Y (2021) "AENeT: an attention-enabled neural architecture for fake news detection using contextual features, Neural Comput Appl Jang SM, Geng T, Li JY, Xia R, Huang CT, Kim H, Tang J (2018) A computational approach for examining the roots and spreading patterns of fake news: evolution tree analysis. Comput Hum Behav 84:103–113

12. Javed MS, Majeed H, Mujtaba H, Beg MO (2021) Fake reviews classification using deep learning ensemble of shallow convolutions. J Comput Soc Sci 4:883–902

13. Jiang T, Li JP, Haq AU, Saboor A, Ali A (2021) A novel stacking approach for accurate detection of fake news. IEEE Access 9:22626–22639

14. Jwa H, Oh D, Park K, Kang J, Lim H (2019) ExBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). Appl Sci 9(19):4062

15. Kaliyar RK, Goswami A, Narang P, Sinha S (2020) FNDNet—a deep convolutional neural network for fake news detection. Cogn Syst Res 61:32–44

16. Kaliyar RK, Goswami A, Narang P (2021a) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 80:11765–11788

17. Kaliyar RK, Goswami A, Narang P (2021b) DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. J Supercomput 77:1015–1037

18. Talwar S, Dhir A, Singh D, Virk GS, Salo J (2020) Sharing of fake news on social media: application of the honeycomb framework and the third-person effect hypothesis. J Retail Consum Serv. https://doi.org/10.1016/j.jretconser.2020.102197

19. Taskin SG, Kucuksille EU, Topal K (2021) Detection of Turkish Fake News in Twitter with Machine Learning Algorithms, Arab J Sci Eng

20. Trueman TE, Kumar A, Narayanasamy P, Vidya J (2021) Attentionbased C-BiLSTM for fake news detection, Appl Soft Comput, 110

21. Umer M, Imtiaz Z, Ullah S, Mehmood A, Choi GS, On B-W (2020) Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access 8:156695–156706

22. Vereshchaka A, Cosimini S, Dong W (2020) Analyzing and distinguishing fake and real news to mitigate the problem of disinformation. Comput Math Organ Theory 26:350–364

23. Verma PK, Agrawal P, Amorim I, Prodan R (2021) WELFake: word embedding over linguistic features for fake news detection. IEEE Trans Comput Soc Syst 8(4):881–893

24. Wang X, Chao F, Yu G, Zhang K (2022) Factors influencing fake news rebuttal acceptance during the COVID-19 pandemic and the moderating effect of cognitive ability. Comput Hum Behav 130:107174

25. Xu K, Wang F, Wang H, Yang B (2020) Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci Technol 25(1):20–27

26. Zhou, X.; Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Comput. Surv. (CSUR) 2020, 53, 1–40. [CrossRef]

27. Zhang, X.; Ghorbani, A.A. An overview of online fake news: Characterization, detection, and discussion. Inf. Process. Manag.2020, 57, 102025. [CrossRef]

28. Hu, L.; Wei, S.; Zhao, Z.; Wu, B. Deep learning for fake news detection: A comprehensive survey. AI Open 2022, 3, 133–155.

29. Athira, A.B.; Kumar, S.M.; Chacko, A.M. A systematic survey on explainable AI applied to fake news detection. Eng. Appl. Artif. Intell. 2023, 122, 106087.

30. 10. Hotho, A.; Nürnberger, A.; Paaß, G. A brief survey of text mining. J. Lang. Technol. Comput. Linguist. 2005, 20, 19–62.

31. Zhou, Z.-H. Machine Learning; Springer Nature: Cham, Switzerland, 2021.

32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]

33. Chowdhary, K.; Chowdhary, K.R. Natural language processing. In Fundamentals of Artificial Intelligence; Springer: New Delhi,India, 2020; pp. 603–649.

34. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. Comput. Intell.Neurosci. 2018, 2018, 7068349. [CrossRef] [PubMed]

35. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989, 1, 541–551. [CrossRef]

36. Elman, J.L. Finding structure in time. Cogn. Sci. 1990, 14, 179–211. [CrossRef]

37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.

Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding.

arXiv 2018, arXiv:1810.04805.

38. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving language understanding by generative pre-training.OpenAI Blog 2018.

39. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners.OpenAI Blog 2019, 1, 9.

40. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A. Languagemodels are few-shot learners. Adv. Neural Inf. Process. Syst. 2020, 33, 1877–1901.