Literature Survey On “ AI-based Detection of Misinformation on Social Media Using Machine Learning and NLP”

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

Authors: [PERSON_58], [PERSON_58], [PERSON_58], [PERSON_58] ([DATE_TIME_36])
This review extensively surveys ML and DL techniques for fake news detection, contrasting traditional classifiers with advanced models like BERT, GANM, and GNN. It demonstrates that transformer-based models consistently outperform traditional approaches, even achieving near-perfect accuracy in some cases. The paper also highlights significant limitations—data quality issues, lack of interpretability, problems with domain generalization, and challenges in real-time deployment. It emphasizes the need for future research focused on explainable models, few-shot learning, lightweight architectures, and ethical considerations for wider adoption.

2. A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection

Authors: [PERSON_58], [PERSON_58], [PERSON_58] (2022)
This study benchmarks classical ML models (SVM, XGBoost), various DL architectures (CNN, BiLSTM), and transformer-based models ([NRP_9], RoBERTabase) using embeddings like GloVe and contextual [PERSON_58] across several datasets. Results indicate that transformer models typically deliver the highest performance, yet no one model excels across all datasets, underscoring the importance of contextual or domain-specific tuning. The research underscores that content-only approaches can still achieve competitive results, guiding future development in scalable fake news detection.

3. AI-Based Fake News Detection Using NLP

Author: [PERSON_58] (2024)
Polu introduces an innovative hybrid framework that integrates transformer-based NLP models (BERT, [PERSON_58]), graph-based credibility scoring using knowledge graphs, and explainable AI tools like LIME and SHAP. Evaluated against real-world datasets (LIAR, FakeNewsNet, BuzzFeedNews), the approach combines semantic, stylistic, and discourse-based features to deliver improved detection accuracy and interpretability. The model also increases trust by providing transparent insights into decision-making—essential for real-world applications. Limitations include adversarial attacks and the need for broader generalization across languages and domains.

■ 4. Artificial Intelligence in the Battle Against Disinformation and Misinformation: A Systematic Review of Challenges and Approaches

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et al. (2025)
This [DATE_TIME_36]-spanning systematic review (2014–2024) synthesizes AI approaches employed against misinformation—from classic NLP and ML to hybrid human-AI methods. It assigns special focus to ethical considerations, the necessity of human oversight, and interdisciplinary solutions. Though comprehensive in scope—highlighting emergent frameworks and cross-sector collaborations—the paper remains theoretical and lacks experimental validation or performance comparisons, signaling a need for future applied research.

5. Big Data ML-Based Fake News Detection Using Distributed Learning

Authors: [PERSON_58], [PERSON_58] (2023)
The study builds a scalable fake news detection pipeline using Apache Spark and a stacked ensemble of classifiers atop large Twitter datasets. It extracts features via N-grams, Hashing TF-IDF, and Count [PERSON_58], producing a significant improvement—achieving a 9.35% higher [US_DRIVER_LICENSE_62]-score compared to baseline methods. This demonstrates the potential of big d for high-throughput fake news detection. However, the system's reliance on centralized Spark clusters raises latency and infrastructure concerns for real-time use.

6. A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et al. (2022)
Proposes a hybrid detection mechanism that merges linguistic analysis (e.g., readability, lexicon, sentiment) with knowledge-based features such as source reputability and fact-checking. It tests ML models like Random Forest and XGBoost, achieving an accuracy of 94.4%—significantly above models using either feature type alone. The study demonstrates the power of combining content analysis with external credibility signals but remains limited to textual data and assumes access to quality fact-checking sources.

7. Neural Networks for Detecting Fake News and Misinformation: An AI-Powered Framework for Securing Digital Media and Social Platforms

Authors: [PERSON_58], [PERSON_58], [PERSON_58], [PERSON_58] ([DATE_TIME_36])
Explores and compares deep learning models—CNN, RNN, LSTM, and transformers like BERT and GPT-3—for real-time fake news detection using datasets such as FakeNewsNet and PHEME. Transformer models achieved precision exceeding 95%, underscoring the relevance of

■ context-aware NLP. Still, the study raises concerns about computational cost, retraining needs due to evolving misinformation, and ethical implications regarding bias and explainability.

8. Conspiracy or Not? A Deep Learning Approach to Spot It on Twitter

Authors: [PERSON_58], [PERSON_58], [PERSON_58], [PERSON_58] (2022)
Introduces BORJIS, a joint conspiracy and sarcasm detection model trained on Twitter data. It consists of a crawler and labeling module, feature vector extraction, and a recurrent neural network-based classifier that integrates sentiment, popularity, and context. Results show a 10%+ improvement over existing techniques, pointing to the value of combining linguistic and contextual metadata for nuanced fake news variants like conspiracies and sarcasm.

9. Motivations, Methods and Metrics of Misinformation Detection: An NLP Perspective

Authors: [PERSON_58], [PERSON_58], [PERSON_58], [PERSON_58] ([DATE_TIME_36])
This NLP-focused survey reviews methods (rule-based, ML, DL), feature representations, related sub-tasks (stance detection, summarization, fact-checking), and evaluation metrics. It encourages hybrid multimodal systems, acknowledges limitations like dataset scarcity and evolving misinformation, and outlines challenges like cross-cultural language variance, emphasizing the open-endedness of the misinformation domain.

10. Explainable Misinformation Detection Across Multiple Social Media
Platforms

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et al. (2023)
Presents a novel, generalizable architecture combining Domain Adversarial Neural Networks (DANN) for cross-platform detection and LIME for explainability. Tested on COVID-19 misinformation datasets (CoAID and [PERSON_58]) across [LOCATION_42], Instagram, [LOCATION_42], articles, the model improves accuracy by 3% and AUC by 9%. LIME provides transparency, enhancing trust in model decisions—addressing key needs in multi-platform misinformation defense.

11. Advanced Detection and Forecasting of Fake News on Social Media
Platforms Using NLP and AI

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et al. (2025)
Combines NLP, sentiment, linguistic cues, and social network analytics to both detect fake news

■ and forecast its spread. Emphasizes predictive modeling alongside detection, advocating for proactive misinformation mitigation. Highlights gaps such as limited dataset diversity, reactive rather than real-time deployment, and the importance of ethical frameworks in AI systems.

12. AI-Assisted Deep NLP-Based Approach for Prediction of Fake News From
Social Media Users

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et
al. (2024)
Builds a four-layered detection architecture—publication, social network, edge, and cloud—that utilizes a CNN-BiLSTM-attention model at the edge for real-time detection and a multi-headed attention scorer in the cloud for credibility validation. Evaluated on Buzzface, FakeNewsNet, and Twitter datasets, it boasts ~99.72% accuracy and 98.33% [US_DRIVER_LICENSE_62] score—exemplifying high-accuracy, context-sensitive user and publisher credibility filtering.

13. Toxic Fake News Detection and Classification for Combating COVID-19
Misinformation

Authors: [PERSON_58], [PERSON_58], [PERSON_58], et al. (2024)
Distinguishes between toxic and non-toxic fake news, particularly in the [PERSON_58] context. It leverages toxicity scoring plus ML models (linear SVM, RF, BERT-SVM/RF), finding linear SVM with toxicity features to be the most effective (92% accuracy, 95% [US_DRIVER_LICENSE_62]). By focusi variations of misinformation, the study supports prioritizing more dangerous content for mitigation strategies.

14. Using and Comparison of Artificial Intelligence Techniques to Detect
Misinformation and Disinformation on Twitter

Authors: [PERSON_58], Funda [PERSON_58] (2024)
Evaluates several AI models—LSTM, SVM, Random Forest, Naïve Bayes, and XGBoost—on a balanced dataset (≈23k fake tweets, ≈21k real tweets). XGBoost outperforms all with 99.82% accuracy and 99.81% [US_DRIVER_LICENSE_62] score, followed closely by SVM and LSTM. Highlights th ensemble methods on high-quality datasets, but scope is limited to text-only Twitter data.

15. Combating Multimodal Fake News on Social Media: Methods, Datasets, and Future Perspective

Authors: [PERSON_58], [PERSON_58] (2022)
Offers a comprehensive survey of DL-based multimodal fake news detection, covering methods

■ (CNN, RvNNs, GANs), pre-trained models (VGGNet, [PERSON_58], [PERSON_58]), transfer learning, be multimodal datasets, and data collection techniques. It underscores the scarcity of rich multimodal datasets, emphasizes the value of visual-text integration, and outlines key research directions like fusion strategies, dataset creation pipelines, and the need for deeper multimodal modeling.

■